Technical & Industrial Sponsors

Proceedings of:
The First International Conference of the IEEE Nigeria Computer Chapter
Theme: The Role of Information and Communication Technology in the Management and Control of Communicable and Non-Communicable Diseases
IEEEnigComputConf'16

(in Collaboration with the Department of Computer Science, Faculty of Communication and Information Sciences, University of Ilorin, Ilorin, Kwara State, Nigeria)

Wednesday, 23rd – Saturday, 26th November, 2016

IEEE Catalog Number: CFP16G60-PRT
Print ISBN: 978-1-5090-6224-9
Online ISBN: 978-1-5090-6223-2

Editors

Prof. Bamidele Oluwade
Dr. A. O. Babatunde
Dr. Nasir Faruk

COPYRIGHT INFORMATION

The First International Conference of the IEEE Nigeria Computer Chapter

IEEE Catalog Number: CFP16G60-PRT
Print ISBN: 978-1-5090-6224-9
Online ISBN: 978-1-5090-6223-2

PREFACE

This book contains the Proceeding of research papers presented at the 1st International Conference of the IEEE Nigeria Computer Chapter (IEEEnigComputConf'16), held between Wednesday, 23rd November, 2016 and Saturday, 26th November, 2016 at the University of Ilorin, Ilorin, Kwara State, Nigeria.

The conference was organized by the IEEE Nigeria Computer Chapter (http://www.ieee.org/go/nigeriacomputerchapter) in collaboration with the Department of Computer Science, Faculty of Communication and Information Sciences, University of Ilorin. The Department of Computer Engineering, Faculty of Engineering and Technology of the same institution also served as a technical co-sponsor.

In all, a total of over sixty (60) papers, including two (2) lead papers were submitted as at the time of going to the press. Apart from Nigeria, submissions were received from such countries as Malaysia, South Africa and Pakistan. The papers were subjected to a referee process with respect to the actual content and the level of originality. The thirty eight (38) papers which appear in this Proceeding were those that substantially met the set acceptance criteria.

I wish to appreciate all who contributed in one form or the other towards making this Proceeding a reality.

Thank you.

Professor Bamidele ('Dele) Oluwade, Ph.D., SMIEEE, FNCS

General Editor,
Proceedings of the 1st International Conference of the IEEE Nigeria Computer Chapter, November 2016.

Chair Central Planning Committee

Professor Bamidele Oluwade

Vice-Chair, Central Planning Committee

Engr. Tunde Y. Salihu

Local Organizing Committee/LOC

- Professor Bamidele Oluwade (Chair)
- Professor J. S. Sadiku
- Dr. Jayeola F. Opadiji (Co-Chair)
- Dr. R.G. Jimoh (Chairman, Venue, Accomodation & Logistics Committee)
- Dr. P. B Shola
- Dr. A. O Babatunde(Chairman, Publications Committee)
- Dr. D. R. Aremu
- Dr. R.O. Oladele
- Dr. A. O. Ameen
- Dr. (Mrs.) T.O. Oladele
- Dr.(Mrs.) O. C. Abikoye
- Dr. (Mrs.) M. A. Mabayoje
- Dr. A. O. Bajeh
- Dr. (Mrs.) F. E. Usman-Hamza
- Dr. J. Joshua
- Dr. A. Y. Abdulrahman
- Dr. A. S. Afolabi
- Engr. S. A. Olatunji
- Mr. K. S. Adewole
- Dr. (Mrs.) A. G. Akintola
- Mr. P. O. Sadiku
- Mr. A. M. Balogun
- Mr. A. O. Balogun
- Mrs. S. A. Salihu
- Miss A. W .Yusuf-Asaju

Editors and Reviewers

- Professor Bamidele Oluwade
- Dr. A. O. Babatunde
- Dr. Nasir Faruk
- Engr. Tunde Y. Salihu
- Dr. J. O Opadiji
- Professor Gloria Chukwudebe
- Sunday Fatokun
- Taye Oladele Aro

Secretariat Registration Committees & Others

- Sharon Ikechi
- Taiwo Tawakalt Olaniyan
- Ayodeji Akinyele
- Adedamola O. Oyeniran
- Oluwafisayo Ayoade

TABLE OF CONTENTS

Operations Research in the Management of Diseases

Joshua O. A. Ayeni
(Professor of Computer Science)
Department of Computer Science,
Redeemer's University,
Ede, Osun State, Nigeria
(Lead Paper)

Abstract **- Diseases are frequently referred to as communicable (infectious or contagious) or non-communicable (chronic). However, some authorities have challenged this grouping over the years. They claim that some non-communicable diseases can be considered communicable, if one considers the causation, whose determinants are considered vectors, of the so-called non-communicable diseases. In fact, Remais B. Justin et al.[3] stated that the two groups converge in low and medium income countries (LMICs) and that a combined strategy is needed in surveillance and disease control. It is therefore, posited in this paper that the management of the two groups in LMICs – communicable and non-communicable need not be separated and so Operations Research methodology should be applied to them alike.**

Key words **- operations research, disease management, health research, low- and middle-income countries**

I. INTRODUCTION

The theme of this conference is "The Role of Information and Communication Technology in Information Management and Control of Communicable and Non-Communicable Diseases". I have decided to speak on the topic: "Operations Research for Management of Diseases".

Communicable diseases have been referred to as diseases that can be transmitted and make people ill. They are caused by pathogenic microorganisms such as bacteria, viruses, parasites or fungi. They can be spread, directly or indirectly, from one person to another. They include: lassa fever, tuberculosis, malaria, HIV/AIDS, Ebola, to name a few. They are also referred to as infectious or contagious diseases. From WHO (World Health Organization) Fact sheets, we read that in 2015, 95 countries and territories had on-going malaria transmission and about 3.2 billion people (almost half of the world's population) were at the risk of malaria. In that year, sub-Saharan Africa had up to 88% of malaria cases and 90% of malaria deaths.

Non-communicable diseases (NCDs) also known as, chronic diseases are not passed from person to person. They are of long duration and generally slow in progression. WHO Fact sheet in January 2015 stated that NCDs killed 38 million people each year and three quarters of them occur in low and middle income countries (including Nigeria). NCDs include: heart attacks and stroke, cancer, asthma and diabetes.

Barret–Connor [17] believes that the distinction between chronic (non-communicable) disease and infectious (communicable) disease is arbitrary and detrimental to epidemiologic study of disease.

M. Ackland et al. [4] stated that chronic non-infectious diseases like cardiovascular diseases are communicable. The reason is that, if one looks at the causal factors of chronic diseases, their (causal factors) determinants should be considered as vectors for the diseases.

Justin V. Remais et al.[3] stated that "the convergence of non-communicable disease (NCD) and infectious disease (ID) in low and middle-income countries (LMICs) presents new challenges and new opportunities to enact responsive changes in policy and research". They opined that a continued strategy is needed in surveillance and disease control.

Operations Research (OR) represents an integrated framework to help make decisions. It is a critically important way to support managerial decision making in health fields essentially in planning, coordinating, training and evaluation functions. The central objective of OR is always to obtain a better understanding of the 'operations' of programmes so that needed improvements can be made. This is achieved through increasing the efficiency, effectiveness and quality of services delivered by providers and the availability, accessibility and acceptability of services desired by users.

In the sections that follow, we will talk about what OR is – its definition, tools and approach. We will also talk about management of disease in health care. We will then give specific applications of OR in health care and conclude with recommendations.

II. OPERATIONS RESEARCH

Operations Research (also known in Europe as Operational Research) originated in the efforts of military planners during World War I. After World War II, the techniques became more widely applied to problems in business, industry and society. The techniques focused on the development of mathematical models that could be used to analyze and optimize complex systems and has become an area of active academic and industrial research to date.

The terms Operation Research and Management Science tend to be used synonymously. A common misconception held by many people is that OR is a collection of mathematical tools. While it is true that it uses a variety of mathematical techniques, it has a broader scope. It is a systematic approach to solving problem which uses one or more analytical tools in the process of analysis. The important thing here is that OR uses a methodology that is objective and clearly articulated and is built around the philosophy that such an approach is superior to one that is based purely on subjectivity and the opinion of "experts" in that it will lead to better and more consistent decisions. In an application of OR, the final decision is taken by somebody who has the knowledge of the problem being solved and can align the results of the analysis to arrive at a sensible decision. It can also be concluded that OR can be translated to mean 'Research' into how 'Operations' can be made better (improved upon).

Tools of Operation Research

OR uses any suitable tools or techniques available. The basic tools include system analysis, probability theory, decision theory, Monte Carlo simulation, stochastic processes, markov decision processes, and queueing models transportation models, Network models, game theory, mathematical programming models, inventory models and other computational techniques such as computational intelligence methods.

OR Approach and Strategy

Wikipedia (Retrieved 19/11/2016) defines strategy as a high level plan to achieve one or more goals under conditions of uncertainty. Bell[2] suggests that the primary impact of a successful strategy is that it creates a competitive advantage that is sustainable over a period of time. Bell & Anderson [1] identified 13 out of 42 private sector (in Edelman Prize finalist articles published between 1990 and 1999) as OR/MS work that leads to sustainable competitive advantage. Thus, OR also addresses strategic problems, i.e., problems that involve managerial decisions.

OR approach to making high level managerial decisions comprises the following sequential steps (see Nikita Dutta [5])

- Establishing objective
- Defining the problem including goals
- Identifying possible alternative courses of action thru:
    - i) Data Collection
    - ii) Model Formulation
    - iii) Solution
- Evaluating alternative courses of action and choosing the best
- Implementation and monitoring

These steps are tied together through a mechanism for continuous feedback.

Brandeau [6] states in her paper that junior scholars/academia often avoid working on practical applications in health because promotion and tenure processes tend to value theoretical studies (i.e., steps i-iii above) more highly than applied studies. She related her experience in using OR to inform and influence decisions in health.

III. MANAGEMENT OF DISEASES

Disease management is a large industry with many vendors and programmes offered through health plans, agencies, associations and employers that offer health insurance. OR-based analysis has potential to improve decision making for real-world health care problems.

In a review paper by William K. Bosu [18], he discussed many areas where OR has been used to assist in the management of infectious (communicable) diseases in developing countries and advocated that since the demarcation between infectious (communicable) diseases and non-communicable diseases were becoming blurred, it is justifiable to integrate the health programmes for the two classes wherever possible. The followings are some considerations in the disease management and delivery of service in developing countries, most of which belong to LMICs.

1. Setting Priorities

A very useful checklist on the process for setting health research priorities has been made available [7]. It outlines what activities and factors should be considered in the preparatory stage, how priorities should be decided on and what actions follow after priorities have been set.

2. Integration of OR results into service delivery and programme management

Academics have integrated a cluster-randomized trial comparing strategies for HIV care into routing health care delivery in Uganda [8].

3. Use of routine data

Routine data have been used to estimate the prevalence of hypertension in workers or other defined population groups [9] or to assess outcomes in a cohort of diabetic patients in primary health care settings after a 3-year follow-up [10]

4. Rapid data Collection and Analysis

Electronically-based STEPS survey or eSTEPS developed by WHO involves the use of personal digital assistants with a suite of software for data collection, data-quality checks and tools for data analysis. It also allows higher-quality data and standardization across studies in different sites.

5. Patient self-monitoring is an essential part of the management of hypertension and diabetes. Rapid tests are being continually developed for the diagnosing of malaria, TB and HIV. Home-based electronic devices are available for measuring blood levels of glucose, cholesterol and triglycerides.

6. Chemoprophylaxis

The principle of chemoprophylaxis aims at prevention of disease when a person comes into contact with infectious disease agent. Aspirin and satins are known to be very effective in secondary prevention.

7. Modeling

Data are needed to feed mathematical models to provide permanent programme information. Aids Impact Model and Spectrum were developed to help estimate the prevalence of HIV in the general population based on surveillance of pregnant women attending prenatal clinics.

IV. SPECIFIC OR APPLICATIONS IN HEALTHCARE

1. Electronic access to diagnostic result systems. The use of information and communication technology (ICT) comes into play here.

2. Vehicle location problem uses iterative optimization algorithm with parameter estimation (see Kim & Lee[11])

3. Queueing model was used to determine number of hospital workers to meet specific service level in a care on demand process (see van Eeden et al.[12]).

4. Dynamic grouping and prioritization (DCP) algorithm was used to identify most appropriate patient groups and prioritises them according to patient- and system- related information, and gives a superior performance to ESI (see Ashour [13]).

5. Simulation-based decision support model was used in the design of an emergency department (ED) to improve patient throughput time goal of arrival to departure under 3 hours for 80% of ED patients (see Chongsun [14]).

6. Computational intelligence relies greatly on historic data (Dounias and Linkens [16]). It is able to make use of adaptive learning algorithms for medical diagnosis (see Amato et al. [15]). Some of these algorithms are evolutionary algorithms and data mining algorithms.

7. The objective of heuristics is to produce a solution in a reasonable time frame that is good enough for solving the problem at hand. They may produce results by themselves or they may be used in conjunction with optimization algorithms to improve their efficiency. Heuristics underlie the whole of AI and computer simulation of thinking as they may be used in situations where there are no known algorithms.

V. CONCLUSION

I summarize this paper with the following recommendations:

1. Academia in Operations Research should extend their research beyond obtaining scientific solutions only. Most real-life problems are more complex than what algorithms and heuristics can solve alone.

2. Governments and communities in low- and middle- income countries (Nigeria included) should endeavor to build capacity to increase volume of research in Operations Research through adequate funding and also provide facilities that would enhance healthy living in the society.

3. Operations Research strategies should be integrated to health care programmes for diseases in LMICs using available international standard tools.

REFERENCES/BIBLIOGRAPHY

1. Bell, P. C., & Anderson, C. K. (2002). In search of strategic operations research/management science [Electronic version]. Interfaces, 32(2), 28-40. Retrieved [19/11/2016], from Cornell University, School of Hospitality Administration site: http://scholarship.sha.cornell.edu/articles/217/

2. Bell, P. C. 1998. Strategic operational research. J. Oper. Res. Soc. 49 381–391.

3. Remais JV, Zeng G, Li G, Tian L, Engelgau MM. Convergence of non-communicable and infectious diseases in low- and middle-income countries. *Int J Epidemiol*. 2012;42(1):221–227.

4. M Ackland, B C K Choi, P Puska. Rethinking the terms non-communicable disease and chronic disease. *J Epidemiol Community Health* 2003;**57**:838-839 doi:10.1136/jech.57.11.838

5. Nikita Dutta, Managerial Decision Making Process (5 Steps). www.economicsdiscussion.net/decision-making/managerial-decision-making-process-5-steps/6099. Retrieved 19/11/2016.

6. Brandeau, M.L. Health Care Manag Sci (2016) 19: 305. doi:10.1007/s10729-015-9328-0

7. Viergever RF, Olifson S, Ghaffar A, Terry RF. A checklist for health research priority setting: nine common themes of good practice. *Health Res Policy Syst*. 2010;8:36.

8. Jaffar S, Amuron B, Birungi J, et al. Integrating research into routine service delivery in an antiretroviral treatment programme: lessons learnt from a cluster randomized trial comparing strategies of HIV care in Jinja, Uganda. *Trop Med Int Health*. 2008;13(6):795–800.

9. Seck SM, Guéye S, Tamba K, Ba I. Prevalence of chronic cardiovascular and metabolic diseases in Senegalese workers: a cross-sectional study, 2010. *Prev Chronic Dis*. 2013;10:110339.

10. Khader A, Ballout G, Shahin Y, et al. Treatment outcomes in a cohort of Palestine refugees with diabetes mellitus followed through use of e-health over 3 years in Jordan. *Trop Med Int Health*. 2014;19(2): 219–223.

11. Kim, S.H. & Lee, Y.H. Health Care Manag Sci (2016) 19: 362. doi:10.1007/s10729-015-9332-4

12. van Eeden, K., Moeke, D. & Bekker, R. Health Care Manag Sci (2016) 19: 227. doi:10.1007/s10729-014-9314-y

13. Ashour, O.M. & Okudan Kremer, G.E. Health Care Manag Sci (2016) 19: 192. doi:10.1007/s10729-014-9311-1

14. Chongsun Oh, April M. Novotny, Pamela L. Carter, Ray K. Ready, Diane D. Campbell, Maureen C. Leckie, Use of a simulation-based decision support tool to improve emergency department throughput. Operations Research for Health Care, Vol. 9. Pp29-39, June 2016.

15. Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. Journal of applied biomedicine, 11(2), 47-58.

16. Dounias, G., & Linkens, D. (2004). Adaptive systems and hybrid computational intelligence in medicine. Artificial Intelligence in Medicine, 32(3), 151-155.

17. Barrett-Connor E, Infectious and chronic disease epidemiology; separate and unequal? Am J Epidemiol 1979; 109: 245-249

18. William K. Bosu, Learning lessons from operational research in infectious diseases: can the same model be used for noncommunicable diseases in developing countries? Advances in Medical Education and Practice 2014:5 469-482

MODELING AND SIMULATION OF INFECTIOUS DISEASES: A ROLE FOR COMPUTATIONAL SCIENTISTS

J. S. Sadiku,
Dean of the Faculty of Communication and Information Sciences,
University of Ilorin, Ilorin, Nigeria.
(Lead Paper)

I. INTRODUCTION

Since the beginning of humanity and human organizations, there have been several threats to its survival by endemic outbreaks of infectious diseases. The Ancient Egyptians were hit by a plaque that ravaged them with ferocity around 1715 B.C. In 1348, the Black Death bubonic plague burst in Europe, and was estimated to have killed over 25 million people in just five years. The pandemic influenza virus of 1918–1919 swept through America, Europe, Asia, and Africa and was also estimated to have killed around 40 million people. Less severe influenza pandemics also re-emerged in 1957 and 1963 which killed up to one and two million people respectively (World Health Organization: http://apps.who.int/iris/handle/10665/68985). In the last decade, people in the world, particularly in Africa continent, have been contending with various emerging and re-emerging epidemics such as AIDS, measles, malaria, polio and tuberculosis that have been claiming the lives of millions of people yearly. According to the UNAIDS report on the global AIDS epidemic, an estimated 34 million people, including 3.4 million children, were living with HIV worldwide at the end of 2010 and not less than 1.8 million people have died as a result of the pandemic.

With rapid technological advancement, we are now able to produce effective vaccines and antiviral drugs, and with a large and intensive research, we are able to continue to design better drugs and vaccines. Nevertheless, humankind will continue to contend or grapple with the challenges of regular outbreaks of communicable and non-communicable diseases and epidemics. Just about two years back, Africa was gripped with the fear of Ebola, which spread like wildfire in West African countries. And we don't forget the recent outbreak of Fikka viruses in Brazil, Avian flu in Southeast Asia and Lassa fever in some parts of Nigeria. The outbreaks of these diseases and their concealed or unpredictable nature created fear and panic in the minds of people in the world.

Thus, one of the major concerns of policy makers and governments in any responsible and responsive society should always be on how to control the outbreaks and transmission of pandemic diseases. This is necessary because there is always a complexity in the disease transmission mechanism, and, of course, this is why there will always be threats of pandemics of infectious diseases.

So, the world at large is expected to work collaboratively together to prevent, control and tackle the outbreaks of infectious diseases of global impact. This is, without doubt, why there has been an intensive worldwide effort in speeding up the developments in the establishment of a global surveillance network for combating pandemics of emergent and re-emergent infectious diseases. Scientists from different fields extending from medicine and molecular biology to computer science and applied mathematics have teamed up for rapid assessment of potentially urgent situations.

Towards this aim, mathematical modeling has been playing an important role in predicting, assessing, and controlling potential outbreaks of pandemics. In other words, there has been an inevitable need to have a better understanding of how infectious diseases spread among a community or a population and how the spread could be quickly controlled. To achieve this understanding, modeling and simulating the contagious dynamics, the impact of numerous variables ranging from the micro host–pathogen level to host-to-host interactions, as well as prevailing ecological, social, economic, and demographic factors across the globe have proved to be utmost useful (Constantinos I. Siettos and Lucia Russo, 2015).

A mathematical model is an explicit mathematical description of the simplified dynamics of a system. A mathematical model is a set of equations, which are the mathematical translation of hypotheses (or assumptions). Mathematical models can enable an extrapolation from current information about the state and progress of an outbreak to predict the future spread as well as possible controls of the spread of the outbreak. So, Mathematical models are now a fundamental element in planning control and mitigation measures against any future epidemic of an infectious disease. Well-parameterized mathematical models allow us to test a variety of possible control strategies in computer simulations before applying them in reality, which might not be possible in epidemiology experiments.

II. THE BEGINNING OF MODELING AND SIMULATION FOR INFECTIOUS DISEASES

The very first epidemiological model was formulated by Daniel Bernoulli in1760 with the aim of evaluating the impact of variolation on human life expectancy. So, Daniel Bernoulli was the first to develop a mathematical model, which was to analyze the mortality due to smallpox in England, and which at that time was one in 14 of the total mortality. Bernoulli used his model to show that inoculation against the virus would increase the life expectancy at birth by about three years. Lambert, in 1772, followed up the work of Bernoulli extending the model by incorporating age-dependent parameters.

At the beginning of the 20th century, efforts at developing modeling for explaining and predicting the spread of infectious diseases re-emerged with the pioneering work of Hamer (1906) and Ross (1911) on measles and malaria, respectively. We can actually say that it was Ross's work that established modern mathematical epidemiology. In his work, Ross addressed the mechanistic a priori modeling approach using a set of equations to approximate the discrete-time dynamics of malaria through the mosquito-borne pathogen transmission.
Following up the work of Ross, Kermack and McKendrick published three seminal papers which founded the deterministic compartmental epidemic modeling. In these papers, they addressed the mass–action incident in disease transmission cycle, suggesting that the probability of infection of a susceptible (virgin from illness) is analogous to the number of its contacts with infected individuals. Thus, in 1927, Kermack and McKendrick derived the celebrated threshold theorem for the purpose of predicting the spreading pattern of an infectious disease. Their model established the critical fraction of susceptibles in the population that must be exceeded if an epidemic is to occur.

Lermack and his team work was followed by the classic work of Bartlett, who examined models and data to expose the factors that determine disease persistence in large populations.

MacDonald extended Ross's model forty years after to explain in more details the transmission process of malaria and to propose methods for eradicating the disease on an operational level. Due to the importance of MacDonald's contribution to the field by exploiting the use of computers, mathematical models for the dynamics and the control of mosquito-transmitted pathogens are known as Ross–MacDonald models (Smith, et.al., 2012).

Enko, in 1889, published a remarkable probabilistic model for describing the epidemic of measles in discrete time. With the use of the model, he evaluated the number of contacts between infectives and susceptibles in the population. Enko's model is the precursor of the famous Reed-Frost chain binomial model introduced by W. H. Frost in 1928. This model assumed that the infection spreads from an infected to a susceptible individual through discrete time Markov chain events. This representation set the basis of contemporary stochastic epidemic modeling. Given the diversity of infectious diseases studied since the middle of the 1950s, an impressive variety of epidemiological models have been developed.

III. PURPOSES OF MODELING AND SIMULATIONS FOR INFECTIOUS DISEASES

The pattern of contacts between individuals is a crucial determinant for the spread of infectious diseases in a population. The topological structure of the contact network of the population, the presence of people with a much larger number of contacts than the mean value , the clustering and presence of well-identified communities of people , and the frequency and duration of contacts, all have important implications for the spread and control of epidemics.
Epidemiology alone cannot provide proper insights into all these parameters of endemics or spread of infectious diseases. Thus, we must know that epidemiology alone cannot successfully control the spread of infectious disease or even control the damage wide spread of infectious diseases can cause. The role of epidemiology is even limited when it comes to cure and interventions. There is a need for simulation and modeling to understand analyze, model, and simulate the dynamics of disease generation and propagation (Koopman, 1996).
Such simulations could serve as dry "laboratories" for a new science of experimental epidemiology in which new population-level interventions could be designed, evaluated, and iteratively refined on simulated epidemics, with tangible benefits for real-world epidemic prevention and control efforts.
Mathematical models can project how infectious diseases progress to show the likely outcome of an epidemic and help inform public health interventions. Models use some basic assumptions and mathematics to find parameters for various infectious diseases and use those parameters to calculate the effects of possible interventions, like mass vaccination programmes.
So, modeling and simulation provide decision support to policy makers on the consequence, mitigation, and response to risks of infectious diseases. To this end, modeling and simulation is at the intersection of public health, biological science, and computer technology.
Mathematical modeling has the potential of probing the complexity of infectious disease dynamics. It also has the potential to feed into public health policy and this is why a wide range of models have been designed and used to support infectious disease control, elimination, and eradication efforts.
So, modeling and simulation for infectious diseases entails using computer to develop a mathematical model of a complex system or process. Through

simulation, we can make a representation of actual communities based on demographic and transportation information. In these kinds of simulated environments, we can introduce infectious agents with certain characteristics and watch how they spread. These kinds of simulation enable us understand how an infectious disease can spread in real life, and then provide us a mechanism of preparation for controlling a pandemic when it breaks out. With modeling and simulation, we can also introduce and evaluate the effectiveness of different interventions such as vaccinations or quarantine.

III. MODELING AND SIMULATION STRATEGIES

Modeling and simulation make use of various strategies. One of such strategies is Statistical simulation. Statistical Simulation and modeling allow for rapid assessment and surveillance through statistical methods. A physician, John Snow, used statistical method to model the spread of cholera in London in 1854. His modeling contributed to the successful eradication of the disease then. A basic regression model had also been proposed and constructed by Serfling to monitor the deaths of influenza based on the seasonal pattern of pneumonia and influenza deaths. Cumulative Sum (CUSUM) has also been a most common used technique for the detection of disease outbreaks. This is achieved by monitoring a cumulative performance measure over time. Through Hidden Markov Models, epidemiological concerns, as to how to infer about the dynamics of a particular infectious disease and forecast its outbreak, can be resolved. For example, we can forecast the evolution of an influenza epidemic by monitoring for example the number of reported cases as recorded through a surveillance network of physicians or in hospital units. Principal Component Analysis (PCA) can also be used to create a single surveillance index that can be used to summarize temporal and spatial trends of malaria, as it had been used in India. (Cohen, 2010)

Mathematical Models can be used to forecast the evolution of a "hypothetical" or on-going epidemic spread. Simulation is also used when the cost of collecting data is prohibitively expensive, or there are a large number of experimental conditions to test. For example, continuum models describe the coarse-grained dynamics of the epidemics in the population. It can be used to study a model for the evolution of the disease as a function of the age and the time since vaccination (Keeling, 1999; Greenhalgh, 1995) or investigate the influence of quarantine or isolation of the infected part of the population (Liu, 1986; Hethcote, 2002). Such models can be explored using powerful analysis techniques for ordinary or partial differential equations or by using *SIR* mass-action model of Kermack and McKendrick (1922).

Stochastic models including discrete and continuous-time individual based Markov-chain models. These are usually individual-level models that relax the hypothesis of the mean field approximations of infinite population and perfect mixing introducing the uniqueness of the individual behavior including multiple heterogeneous characteristics.

Lekone and Finkenstadt (2006) had used a stochastic SEIR model to simulate the dynamics of Ebola outbreak in the Democratic Republic of Congo in 1995. Bishai and his team (2011) used a stochastic SIR model with age structure and two additional states (compartments) to describe heterogeneity in vaccination. The authors combined the epidemic model with an economic model incorporating the costs of the control disease policies to study the cost effectiveness of supplemental immunization activities for measles in Uganda. Wang and his group (2012) developed a stochastic model within the SIR concept to simulate and better understand the multi-periodic patterns in outbreaks of Avian flu in North America.

Generally, there are three distinct categories of modeling that could be used to study and understand infectious diseases and their propagation, spread and control. The three categories are Agent Based Models, Network Models, and Digital Models. Agent-based modeling represents the state-of-the-art for reasoning about and simulating complex epidemic systems. It simulates how individuals in a community created on the computer interact with the environment and with one another in order to predict how a disease could be spread over a space and time. These take into account details such as the transportation infrastructure of the simulated area, the mobility of the population, demographics, and epidemiological aspects such as the evolution of the disease within a host and transmission between hosts. Public-health epidemiologists, researchers, and policy makers are turning to these detailed models for reasons of ethics, cost, timeliness, and appropriateness. In epidemic systems, testing experimental conditions would put the safety of people at risk, creating an ethical problem.

V. CONCLUSION

By clarifying rigorously the assumptions entrenched in a model, the variables, and the parameters of a disease, mathematical modeling allows understanding the observed spread of diseases in space and time. Epidemiological model further provides important conceptual results including the basic reproduction number, the threshold effects, and the herd immunity. For evident ethical and practical reasons, experiments in public health are often impossible to perform and mathematical models thus appear as a cheap and efficient way to explore and test hypotheses. This is, for example, of particular practical utility in the design of vaccination policies.

Evaluating the efficacy of control strategies through controlled experiments is usually practically infeasible and unethical, and this is why mathematical modeling are used to provide a powerful tool for conducting such experiments. The advent of computers has enabled us design complex models that can take into account the homogenous nature of diseases and population can now be constructed and studied with simulation and dynamic

analysis software. Thus, computer simulation can now play an important role in the study of infectious diseases and how to control their spread.

To this end, the computer scientists must rise to the occasion of predicting the pattern of likely outbreaks of pandemic diseases in our society with the aim of controlling their spread and mitigating their impact on the lives of the people. As people rely on epidemiologists to save them from pandemic diseases that can gradually or quickly wipe out our population, the people also depend on computer computational technologists, mathematicians and statisticians to use modeling and simulations to predict and prepare us for any emergent outbreaks that could be destructive.

It is therefore recommend that we should all come together to establish a Nigeria Center for Discrete Mathematics and Theoretic Computer Science to focus on computational and mathematical epidemiology for the purpose of identifying and exploring methods in mathematical science not yet widely used in studying epidemiological problems. With this, we would have been contributing our own quota to preventing and eradicating infectious diseases in our society. It is the belief that the different papers that will be presented at this conference will direct us and strengthen our resolve to make our world and society better for living.

Thank you for your attention.

REFERENCES

1. Bartlett, M. S. (1960). The critical community size for measles in the United States. *J R Stat Soc*. **123**:37–44.
2. Bernoulli, D (1760). Essai d'une nouvelle analyse de la mortalité cause par la petite vérole et des avantages de l'inoculum pour la prévenir. *Mém Math Phys Acad Roy Sci Paris*;1–45.
3. Bishai, D, Johns, B, Nair D, Nabyonga-Orem, J, Fiona-Makmot, B, Simons, E. (2011). The cost-effectiveness of supplementary immunization activities for measles: a stochastic model for Uganda. J Infect Dis. 2011;204(*Suppl* 1):S107–15. doi: 10.1093/infdis/jir131
4. Cohen, A.A,, Dhingra, N,, Jotkar R.M,, Rodriguez ,P.S, ,Sharma, V.P,,Jha P. (2010). The Summary Index of Malaria Surveillance (SIMS): a stable index of malaria within India. *Popul Health Metr*. 2010;8:1. doi: 10.1186/1478-7954-8-1.
5. Dietz, K. (1988). The first epidemic model: A historical note on P.D. En'ko. Aust J. *Stat.* 1988;30A:56–65. doi: 10.1111/j.1467-842X.1988.tb00464.x
6. En'ko, P. D. (1889). On the course of epidemics of some infectious diseases. Vrach. St. Petersburg 1889; 1008-1010. 1039-1042, 1061-1063
7. Frost, W.H. (1976). Some conceptions of epidemics in general by Wade Hampton Frost. A.M. J. *Epidemiol*. 1976;103:141–51.
8. Greenhalgh, D,, Das, R. (1995). Modelling epidemics with variable contact rates. Theor Popul Biol. 1995;47:129–79. doi: 10.1006/tpbi.1995.1006
9. Hamer W. H. (1906). Epidemic disease in England. *Lancet* ;**i**:733–9.
10. Hethcote, H.W, Zhien, M, Shengbing L. (2002). Effects of quarantine in six endemic models for infectious diseases. Math Biosci. 2002;180:141–60. doi: 10.1016/S0025-5564(02)00111-6.
11. Kermack, W.O., McKendrick, A.G. (1927). A contribution to the mathematical theory of epidemics. *Proc R Soc Lond* ;**A115**: 700–21.
12. Kermack, W.O, McKendrick, A.G. (1932). Contributions to the mathematical theory of epidemics, part II. *Proc R Soc* Lond.;138:55–83. doi: 10.1098/rspa.1932.0171.
13. Keeling, M. J. (1999). The effects of local spatial structure on epidemiological invasions. *Proc Biol Sci*. 1999;266:859–67. doi: 10.1098/rspb.1999.0716.
14. Lambert, J. H. (1772). Die Toedlichkeit der Kinderblattern. Beytrage zum Gebrauche der Mathematik und deren Anwendung. Buchhandlung der Realschule.;3:568.
15. Lekone P.E., Finkenstädt B.F. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics.* 2006;62:1170–7. doi: 10.1111/j.1541-0420.2006.00609.x
16. Liu W.M, Levin S.A, Iwasa Y. (1986). Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. *J Math Biol*. 1986;23:187–204. doi: 10.1007/BF00276956
17. Ross R. (1911). *The Prevention of Malaria*. Murray, London.
18. Smith D.L, Battle K.E, Hay SI, Barker C.M, Scott T.W, McKenzie F.E. (2012). Ross, macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. PLoS Pathog. 2012;8:e1002588. doi: 10.1371/journal.ppat.1002588.
19. Wang R-H, Jin Z, Liu Q-X, van de Koppel J, Alonso D. A simple stochastic model with environmental transmission explains multi-year periodicity in outbreaks of avian flu. PLoS One. 2012;7:e28873. doi: 10.1371/journal.pone.0028873.

Expert System in Malaria Diagnosis: Research Issues, Tools, And Applications

Joseph Bamidele AWOTUNDE
Computer Science Department
University of Ilorin
Ilorin, Kwara State, Nigeria
awotunde.jb@unilorin.edu.ng

Opeyemi Emmanuel MATILUKO
Center for Systems and Information Services
Landmark University
Omu-Aran, Kwara State, Nigeria.
opmat01@yahoo.com, matiluko.opeyemi@lmu.edu.ng

***Abstract*— In this paper, the author identified a number of advanced research discussions, reviewed some tools and methodologies currently used in an expert system designed for Malaria control, and presented a case study for diagnosis and management of malaria as discussed in this paper. The need to review and analyze different methods that has been used is very vital so as to help anticipated researchers. This will help the researchers understand what have been done since some literatures have been reviewed and this will motivate them to work under the domain and with other diseases since the methods have been analyzed. The analysis is based on the different approaches, the result and conclusions of each work also presented. The author evaluated some selected studies stated in the literature on techniques used in diagnosis and treatment of malaria fever. The studies identify methods that are mostly implored in carrying out research in this area. Findings from this evaluation would go a long way in enlightening and bridging the identified gaps for an improvement of the present study and further studies in this subject area.**

Keywords*— *Expert system, Malaria, Diagnosis, Research, Application, Plasmodium spp

I. INTRODUCTION

Expert system (ES) is the branch of soft computing in computer science that deals with the representation of human proficiency using computers technology. The ES is artificial intelligence that uses computer to solve demanding complications via professional's handle and competency in medicine sciences and many others fields. The structures of ESs are numerous according to diverse technologies used in accomplishing their tasks. However, they derived their terminologies and technologies based on the first set of ES, such as [1]. The ability for ES to diagnose diseases in an interactive setting is an important part of the system. Patient expresses their symptoms and signs through the knowledge base system since diagnosis involves the effective nature of ailment by considering patients signs and symptoms. The system which used knowledge base system in diagnosing diseases is called expert system. The introduction of artificial intelligence (AI) software systems have technologically advanced to support in the process of medical diagnosis. They have also been used as decision support systems for medical scientist. Most present existing automated disease diagnosis systems are either standalone, mobile or web-based systems. Researchers' definition for the AI field is "the study and design of intelligent agents," and the intelligent agents are system that perceives its settings and takes steps that exploit its chances of access [2].

Many algorithms and technologies have been developed using soft computing algorithms such as fuzzy Logic, Genetic Algorithms and Artificial Neutral Network to find solution to the problem of accuracy and precision in the field of medical sciences [3], [4], [5], [6], [7], and [8]. AI has been used in different ways in the health industry, the use of computers as an Expert System to speed up diagnosis of diseases and to confirm it with medical doctors' diagnosis, it provided quick and precise guidance in medical sciences on the diagnosis of related diseases and in most cases prescribe drugs to be used for patient, behave as a counsellor and create awareness to patient on certain diseases [9]. Today, medical diagnostic processes have been automated with computer-related technologies and its use increases almost every day. The basis of these systems, based on the principles of AI, are not just to diagnose patients but also in most cases prescribe drugs and treatment based on such.

Malaria is as old as human being because it is an ancient disease, which causes social, economic and health burden amongst people in the world [9]. The disease is common in the tropical countries and the disease has been in existence for hundreds of years, malaria have remained a prominent public health challenge among many countries. WHO declared malaria endemic in 109 countries in year 2008 with 243 million malaria cases recorded and million deaths of the endemic which are mainly children under 5 years of age (WHO).

Mostly ecological factors and many others help in the spread of malaria, which includes season, climate etc., it is worthy of note that it is not limited to these. The *Anopheles gambiae* mosquito one of the seven classes of mosquitoes is the most effective vector that transmit malaria. They are common in Africa and temperate regions since they cannot survive frosty climate [10]. *Plasmodium falciparum*, is a parasite and protozoan carriers of female Anopheles

mosquitoes that are the main instrumental parasite of malaria. Temperature helps the transmission sequence of *P. falciparum* in many ways, but the effects on the duration of the sporogonic cycle of the parasite and the vector survival are particularly significant. "At temperatures below about 22°C the decisive factor is the number of mosquitoes surviving the parasite's maturation period which takes 55 days at 18°C" [11] and ceases at around 16°C. After 55 days the proportion of a legion of mosquitoes that survives is only 0.003[12].

The disease happen and occur within the geographical boundaries of the affected regions, the problems become worrisome if the conditions between human, parasite and mosquito vector are symmetry then the affected population are disturbed. This problem has increased the number of incidence in Africans with over 124 million live in this types of environment and about 12 million cases of malaria were recorded and about 310,000 deaths annually because this epidemics [13].

Research have shown that malaria remains one of the major public health challenge in Africa [14]. The efforts that have being made in controlling malaria spread and transmissions within and between have not yield the desire results. Utzinger, Yesim, and Burton reported that the malaria monthly incidence rates and vector densities were normally used for surveillance environment management strategies and for adaptive tuning, this have shown a high level of performance in the areas where this formulation have been used [15]. As a result of these methods used, within 3-5 years, the malaria related illness and death rates were reduced by 70-95% [15]. The same authors in recent work, showed that malaria control programmes using environmental management were effective in reducing illness and death [15].

The crucial aim of the work is to offer a summary of techniques for diagnosis of malaria fever using Medical Diagnostic Decision System method, with an overview of their uses, and to offer a framework and future guidelines to researchers who are new and willing to work in this direction.

This review is organized as follows: section 2 introduces various issues relevant to the expert system in malaria. Section 3 reviews tools, and overall methods, that are suitable for conducting research in using expert system in malaria fever and its diagnosis. Section 4 concludes the paper by recommended possible future directions for research in malaria using expert system.

II. RESEARCH ISSUES

The study looks into research issues using modeling, data collection, diagnosis, treatment in malaria fever. It also studies various research issues and challenges facing potential applications in expert system techniques.

A. Modeling in Malaria Disease

The first task and challenge in designing expert system in malaria is the development of an appropriate model. Someone have to query and seek the model that is best appropriate and describes the symptoms, signs and properties of the epidemic. A robust and good mathematical model can help in achieving deeper insights into the interactions existing between patients, symptoms/signs, diagnosis, treatment and drugs in the malaria epidemic.

A well-developed mathematical model help researchers to understands various concepts of the disease at abstract level, and make the definition of terms to be more understandable without any misconception. The developed models can be used in designing a robust expert system and can be used in tackling several other challenges associated with design of expert system. Also, it can be useful in generating an AI dataset, modification of parameters to simulate a special scenario, which can be used to compare different algorithms and studies. Such formulated model will help to study the peculiarities of the use of ES by infer latent patterns and structures that would explain certain phenomena like diagnosis and treatment of malaria, etc.

Medical Diagnostic Decision System's (MDDS) purpose is to support physicians in clinical decision making [16]. People are known for making vital errors when it comes to decision making in medicine, because of high complication of medical issues coupled with cerebral limitations. Coming up with decision-making can be complex in medicine because it required enormous volume of knowledge to even solve seemingly simple problems [17]. Normally, doctors are to remember and apply quite a number of knowledge of many array of individuals' such as disease presentations, indicative parameters, drug mixtures and recommendation for a given diagnosed ailment [18]. However, the physician's cognitive abilities are restricted owing to factors like multi-tasking, restricted reasoning and memory capacity [19], [20]. Consequently, it is difficult for an unaided Medical Practitioner (MP) to make right verdicts all the time [21]. Unfortunately, the increasing level of information generation by medical advances has worsened the MP's task [22]: [21]. Individuals have a "limited channel capacity" [19], due to which they are incapable to process enormous sum of data. These confines result in an estimated 30% illness and death [23]; [24], which is avoidable.

The commonly used models in malaria epidemic are: Analytic Hierarchy Process e.g. [25]; [26]; [27], Fuzzy logic e.g. [9]; [28]; [29]; [30]; [31]; [32]; [33]; [34]; [35]; [36], Neutral Networks e.g. [37]; [38], Decision Tree Analysis e.g. [39]; [40]; [41]; [42] and Automated Image Processing Method e.g. [43]; [44]; [45]; [46]; [47]; [48]; [49]; [50].

B. Data Collection

The collection of an organized electronic health information about a patient or population is term Electronic Health Record (EHR) [51]. This is represented and recorded in digital format, which can be shared across different health care locations. In most cases, the sharing occur through network-connected, enterprise-wide information systems and other information networks or exchanges [52]. Keeping of records is an essential part of health care delivery since without it, planning, scheduling and keeping tracks of patient becomes difficult [53]. Treatment of patients depends mostly on previous records so as to know which drugs to prescribes and administer and also health center base their decisions on records for them to predict or know which drugs to stock, drugs to administer and which services to prioritize. Giving proper attention to health records management cannot be over-emphasized because it make right information to be available at the right time and at the right place. The innovation of storing the patient's information electronically was brought about by making available comprehensive medical information when needed [54]. The main catalyst for electronic health record (EHR) is to improve on patients' medical care.

Data collection is very vital in the treatment of any patient in a hospital as it is used to know the rudimentary history about the patient to be treated. Also, knowing the history of any patient before recommending any treatment for such patient is important because patients differ from each other and while one may like taking drugs the other may prefer injection. Most MDDS stores data from the initial consultation; and the clinician input data on the patient's progress from time to time, such as further parasite counts and temperature readings. The system will then take this data and make it accessible for medical diagnosis or medical research purposes. The data is stored on the server which makes it possible for interested parties to collect, or request for the data. The data can also be mailed to researchers at any location. Most systems store/records nearly all of the information it receives, and also all conclusions that it derives on a permanent database on the server.

C. Malaria Diagnosis

Medical diagnostics is based on different methods of research and determination of diseases and their severity with the purpose to aid, select and apply necessary treatment, and avert the development of complications and repetitive diseases. Diagnostic techniques are methods, which involve interaction between the medical consultant and the patient in form of "question and answer", this made it a good candidate for automation [55].

Diagnosis is the finding and recording of symptoms and signs of abnormal condition that affects a specific patient, this can be done using clinical data or lacerations. The final diagnosis must agrees with the disease that affects the patient, before declaring that the diagnostic process is correct; otherwise, the diagnosis is declared not correct and a misdiagnosis occurred. The diagnostic procedure used the mathematical model design and this is based on disease models stored in the computer knowledge-based. The models must take into consideration the following: the name of disease with the cause, pathogenesis, lesion, pathophysiology, clinical data, syndromes, clinical presentation and complications.

Proper treatment of any aliment need speedy and accurate diagnosis and this is very critical to the effective control and management of malaria fever. Globally the effect of malaria have encouraged the development of an effective diagnostic strategies for proper treatment of malaria in resource-limited areas where malaria is a significant burden on society and in developed countries where malaria diagnostic expertise is often lacking [56]; [57]; [58]. For proper malaria diagnosis, identifying malaria parasites or antigens/products in patient blood is very essential. The statement look simple, but effective diagnostic is subject to many factors.

Malaria parasitemia diagnostic identification and interpretation in a diagnostic test can be influenced by the following: by the five malaria species; different stages of erythrocytic schizogony, endemicity of different species, interrelation between levels of transmission, population movement, parasitemia, immunity, signs and symptoms; drug resistance, the problems of recurrent malaria, persisting viable or non-viable parasitemia, sequestration of the parasites in the deeper tissues, and the use of chemoprophylaxis or even presumptive treatment on the basis of clinical diagnosis [56]. Traditional methods of malaria diagnostic remain problematic. The modern methods of malaria diagnosis have been developed and introduced to overcome the limitation, inaccuracy and impression of clinical diagnosis.

Medical doctors still preferred the used of traditional method of clinical diagnosis of malaria, the method is the least expensive and thus, most widely practiced. The method based on the patients' signs and symptoms, and the findings during medical or laboratory examination. Malaria fever can be detected, nonspecific and inconsistent in its earliest stages of symptoms/signs. These symptoms includes fever, headache, weakness, myalgia, chills, dizziness, abdominal pain, diarrhea, nausea, vomiting, anorexia, and pruritus, which can leads to other ailment if not properly cared for [59]. The traditional used of clinical diagnosis for the epidemic is still problematic and challenging because of the non-specific nature of the signs and symptoms, which has considerable similarity with other common ailment like Laser fever, Ebola virus and many others ailments, including potentially life-threatening diseases like common viral or bacterial infections, and other febrile illnesses [56]. The relationship existing between malaria symptoms and other tropical diseases have created diagnostic problems and specificity. These can promote indiscriminate use of anti-malaria drugs, which will compromise effects and quality of care for patients with non-malarial fevers in prevalent areas [60]; [61]; [56].

Prompt response and effective diagnosis of malaria reduces both complications and death in patients. The difference in clinical diagnosis of malaria from other tropical infections based on patients' signs and symptoms, laboratory examination or physicians' findings, may be very difficult to identify [56] that is why medical diagnostic decision system is a prompt intervention and are urgently needed for diagnosis and treatment of malaria fever. The paper review the intervention brought by the expert system looking into currently available diagnostic methods for malaria using expert system in different models. In other to reduce, manage or limit the effects of malaria, prompt and accurate diagnosis is paramount.

D. Treatment for Malaria

After being successfully diagnosed, the next step for a patient is treatment. Due to malaria similarity with others tropical ailments medical practitioners prefer that treatment for malaria should not be started until proper diagnosis is done and has been established by laboratory investigations. "Presumptive treatment" without the benefit of laboratory confirmation is not encouraging and should not be adopted expect for extreme circumstances (strong clinical suspicion, severe disease, impossibility of obtaining prompt laboratory diagnosis) [55].

After laboratory examination and proper clinical diagnosis of malaria has been made, appropriate anti-malaria treatment must be initiated immediately. Before administering treatment on a patient the following three main factors must be used as a guide:

- The infecting *Plasmodium* species
- The clinical status of the patient
- The drug vulnerability of the infecting parasites as determined by the geographic area where the infection was acquired and the previous use of anti-malaria medicines

The parasitological status of a patient should be properly monitored after the initiation of treatment. In infections with *P. falciparum* or suspected chloroquine-resistant *P. vivax*, blood smears should be made to confirm adequate parasitological response to treatment (decrease in parasite density) [55]. Most of the MDDS under study did not include treatment in their proposed system; this may be because the system cannot include laboratory diagnosis which is very paramount in treating malaria fever and helps to know the types of drugs to recommend to the patient.

III. DIAGNOSIS OF MALARIA: A REVIEW

The method used in this appraisal of work reported in the literature on the techniques used in diagnosis and management of malaria fever was done through the study of journals and publications. The study also includes books chapters, dissertation, working papers and conference papers. Publications considered in this review are recent ones and findings from this appraisal would go a long way in coming up with a robust method to further the research in the development of clinical decision support system for the treatment of malaria and other related diseases.

A. Discussion

Medical Diagnosis especially for diseases that present itself with multiple symptoms, which can be similar to symptoms of other diseases has been found to be daunting.

This section presents a brief summary of results obtained from the research and the discussions of same.

[27] developed a model for subjective evaluation of malaria symptoms using Analytical Hierarchy Process (AHP), techniques developed by Thomas L. Saaty in the 1970s. When tested with sample data from Malaria patients, a confidence level of about 7% above the manual method of diagnosis was achieved. The authors however suggested that the system be subjected to clinic test to aid in deployment of a full scale Malaria Diagnostic system.

[25] The study considered data retrieved from 30 malaria patients and used for diagnosis in two different medical diagnostic expert system built on Analytic Hierarchy Process (AHP) and fuzzy logic respectively. The result showed that results from the system based on fuzzy logic has little more accuracy than that based on AHP however this result is non-statistically represented.

[28] The authors built a Clinical/medical diagnostic system using Visual Prolog programming language based on fuzzy logic. The system was designed to ensure accuracy and precision of laboratory diagnosis of Malaria through the system rather than the traditional method.

[31] The authors concentrated on Fuzzy Expert System for Malaria detection and the system showed high potential with high detection accuracy, which in turn is able to reduce the time and effort put into traditional laboratory diagnosis of malaria without compromising the result accuracy and integrity.

[62] The authors presented an architectural framework for a Coactive Neuro-Fuzzy Expert System model, which is a model design for easy diagnosis of Malaria based on a collection of symptoms. The model was built based on established symptoms of Malaria as reported by patients and an hybrid learning algorithm was used, which was made up of supervised learning (e.g. Back Propagation Network) and unsupervised learning (e.g. Kohonen Self Organizing Feature Map) to make it adaptive in handling cases that have not been predefined in the knowledge base.

[63] The study used crisp input, which were fuzzified using inter-valued and triangular-shaped membership function and weighted average to defuzzify the inference engine output to develop a mobile device based fuzzy expert system for the diagnosis of Malaria. The system leveraged on the gross availability of mobile devices with almost all patient and this was found to be effective in diagnosis of Malaria.

[40] Developed FESM, a system for the management of Malaria based on Fuzzy Logic, which is decision support platform that can be employed in endemic regions to assist Researchers and Healthcare givers. The system was tested with a sample size of 35 randomly selected patient. The system used a triangular-shaped membership function for fuzzification of scalar inputs, root sum square (RSS) for inference and the center of gravity method for defuzzication.

[42] proposed a clinical conventionally-based decision support system (PBDSS) to aid in treatment of malaria infected patients based on the severity level of the infection.

[64] Developed an expert system from the combination of two systems to give a diagnostic tool that can provide prompt diagnosis and this was employed in the Niger-Delta region of Nigeria. The expert system is a combination of Disease-Oriented Health Information System and Action-Oriented Integrated Management of Child Illness.

[65] Studied the possibility of replacing clinical judgment and diagnosis of malaria by the developed decision tree model of diagnosis. The model was in detecting Plasmodium *falciparum* in young febrile children of ages less than 5 years.

[67] Developed an algorithm for effective malaria diagnosis in pediatrics. This can be used by health worker even when at outpost or on outreaches. The algorithm was tested by the Medical Research Council Clinic, Basse, Gambia and it yielded encouraging result in the test sample of about 518 children.

[35] developed a decision support system for mosquito borne disease diagnosis, where medical experts are not easily available i.e. rural areas (or) remote areas. It is a symptom based decision support system. Which will be very useful in the diagnosis of malaria. MATLAB's GUI feature was used in designing the proposed system with the implementation of fuzzy logic.

[37] The authors investigated Malaria detection and diagnosis using two techniques. The first technique used passive case detection on individuals with symptoms of malaria from a community in the Brazilian Amazon by comparing the performance of RDT Optimal-IT, nested PCR and light microscopy. The second technique did an active case detection of 380 individuals from a riverine community in Rondônia, Brazil suffering from malaria without known symptoms. Epidemiological data was also used to compare the performance of nested PCR, microscopy and MalDANN, an expert malaria detection system based on artificial neural networks. This showed that although RDT outperformed microscopy in diagnosis of malaria in patients with low parasitaemia, it is however unable to fully discriminate the Plasmodium species in 12 cases with mixed infections (Plasmodium vivax + Plasmodium falciparum). It was also discovered that the microscopy approached lagged behind in the detection of cases without known malaria symptoms (61.25% of correct diagnoses).

[38] Developed Automatic Diagnosis of Malaria Parasite using Neural Network and Support Vector Machine. The motivation of the research was premised on the high mortality rate of about 300million per annum as a result of malaria. It is believed that early and faster identification of the malaria parasite can help reduce this figure.

[42] Developed a Decision Support System, which is driven by medical and clinical conventions, used in active treatment of Malaria. The motivations for the research are: The medical field has become overwhelmed by large volume of data to manage, resulting into variations in treatment processes, which sometimes lowers quality of service. Malaria has continued to be a global scourge, killing several millions of people annually, larger percentage of which are infants, young children and pregnant mothers in remote or rural areas of Africa. These individuals in most cases have little or no access to standard health facilities and personnel. Additionally, Protocol non-compliance, inadequate knowledge and expertise are all responsible for these millions of death.

[68] Used machine learning techniques in designing an online diagnostic solution. Their studies shows that considering five distinguishable cases of malaria for classification, a 100% accuracy value was attained for the training set while 94% was attained for the testing dataset. Their studies was based on rough set approach, a machine learning paradigm been applied to set of malaria symptoms labelled appropriately to define rules for determining malaria severity. However, the use of platforms like this is often limited by the I.T. skills of the end user.

[49] Developed a system using image processing for identifying and analyzing malaria parasite. The system was developed mainly because the manual Microscopy as a method of diagnosing malaria is time consuming and prone to human error even in experienced hands. The method adopted was Support Vector Machine (SVM) for classification of images based on malaria parasite infestation. The system uses statistical learning theory, which tends to produce predictive function from dataset in the problem domain for informative decision making. The automated system was designed with the aim of reducing or possibly eliminate human factor while providing an optimum tool to accomplish this.

[69] Developed a rule based decision support tool (CLIPS) and a constrained Bayesian model were developed in Netica for comparative analysis of the value of probabilistic frameworks. CLIPS was designed to improve on the time expended in attaining a concise and reliable diagnosis of various malaria parasite species. The designed Bayesian model and CLIPS utilized a few assumption in achieving its aim.

[39] Developed a cell phone-mounted light magnifying device and showed its potential for clinical. P. *falciparum*-infected and sickle red blood cells in brightfield and M. tuberculosis-infected phlegm image samples were taken and tested in fluorescence. This produced results in blood cell and microorganism forms identification. The High resolution digital images were also leveraged on in obtaining counts of bacillus in the tuberculosis phlegm using image analysis tools. It is the opinion of the authors that this can be adopted

worldwide because mobile phones are readily available and other components are less expensive. This is believed to be able to meet the need of medical laboratories in remote areas of developing countries were cost of diagnostics equipment is a barrier.

[70] Designed a software, MalariaCount, that was able to generate parasitemia from images of Giemsa-stained blood smears. This was then applied to typical vitro cultures of Plasmodium *falciparum* and those treated by drugs to determine the potency of MalariaCount. This yielded a close similarity in values obtained from the software and the manual approach of parasitemia count to proof that MalariaCount can effectively be utilized in research laboratories requiring prompt parasitemia count test result in large scale.

[47] Designed a novel technique that used a database of 475 images in identifying malaria parasite via morphological operations, histogram equalization, connected components analysis and thresholding to determine the estimated malaria parasite density. This created an optimized solution in terms of reduced processing time (about 2 seconds per images) and optimum average accuracy of about 96.46%.

[46] Developed a novel image transforming algorithm, in form of an expert system, to dependably identify malaria parasite from Plasmodium *falciparum species* in thin smears of Giemsa stained blood sample. The algorithm adopted employed the same microscopy diagnosis of malaria. To achieve this, digital camera was connected to a light microscope via which digital images were acquired and the same images were converted to grayscale thereby reducing the variability of the images. The proposed algorithm was tested using malaria samples from Eijkman Institute of Molecular Biology, Indonesia.

IV. SUMMARY AND CONCLUSION

The foremost objective of biomedical informatics and Telemedicine is the provisioning of automated decision supports to aid activities of medical practitioners. The models for emerging accurate medical diagnostic system algorithms are essential to achieve this aim. The importance of Medical Diagnostic Systems cannot be over-emphasized today, in view of the ever increasing medical science information eruption. Implementation of Electronic Medical Record (EMR) systems at health-care centers will somewhat necessitate the adoption of Medical Diagnostic System applications. From the reviewed works, researchers working on Medical Diagnostic Decision System for Malaria diagnosis and treatment had approached it using Analytic Hierarchy Process, Fuzzy logic, Neural Network and Automated Image Processing Method. The researchers working on these approaches have made use of different techniques like Symptomatic, Treatment based on laboratory results, imaging and data mining.

The synthesis of Medical Science and expert system in managing malaria fever as a result problems associated with this disease was also established. These models are either based on available symptoms or images of the malaria parasites. The models are expected to be of immense assistance in both urban and rural areas of the affected regions. However, adequate care must be taken while developing these computer based systems. Accuracy and reliability of such systems must be thoroughly evaluated. It is observed from the reviewed works that most of these Computer Based Systems for malaria diagnosis are based on a single predictive models, most provide diagnosis without therapy and vice versa. In addition, most researchers failed to evaluate the detection rate (accuracy) of the designed systems.

The areas of improvements that were identified by the author are:

Inclusion of EMR System in designing MDS in malaria diagnosis will helps in giving precise, appropriate and sufficient prescription (drugs) to patient that use the system since such patient medical history is electronic available.

Computer Based Systems for malaria diagnosis should move from a single predictive model by providing diagnosis with treatment and vice versa.

Accurate Assessments of malaria diagnosis systems in clinical situation are indispensable to foster medical specialists' application of these MDS supports.

Comparative analysis of various forms of modeling techniques is expedient to aid better understanding of these various techniques, their strength and weakness. This also serves as guide to future researchers in making informative decision on the choice of existing model or technique that will favourably apply to a given medical decision problem

Accessibly and simplicity of these systems must also be put into consideration by future researchers. Researchers can attempt to make these systems web based or mobile application based, especially those working in symptomatic environment so that many people can have access to it and the purpose of creating such system will be met.

Future researchers can work on improved accuracy of malaria diagnostic system by modeling imprecise, inaccurate, temporal and historical data, work further on improving inference algorithms, and also find better avenues of collating globally diversified patient information without violating privacy issues as well as optimizing existing clinical microscopy algorithms. Diagnostics systems for medical purposes would be better suited if access to untagged or anonymous medical dataset are made ready available and the development of open-source data mining software/tools are improved on.

References

[1] Mycn Wu, H., (2008). "Research on English Teaching Reading Based on Expert System" IEEE International Conference on Computer Science and Software Engineering, pp. 450-453.

[2] Jaryani, F., Sahibudin, S., Ibrahim, S., Rahman, N.A. and Daruis, R. (2011). "Intelligent Reflective E-Portfolio Framework Based on Artificial Intelligence Expert Systems Techniques", IEEE 978-1-61284-840-2/11, pp. 214-216.

[3] Ledley, R. S., Lusted, L. B., (1959). Reasoning Foundations of Medical Diagnosis. *Science*, 3. July 1959, Vol. 130, Nr. 3366, pp. 9-21.

[4] Lipkin, M., Hardy, J. D., (1958). Mechanical Correlation of Data in Differential Diagnosis of Hematological Diseases. *Journal of the American Medical Association*, 166 (1958), pp. 113-125.

[5] Ledley, R. S., Lusted, L. B., (1962). Medical Diagnosis and Modern Decision Making. In: Bellman, R. (ed.): *Mathematical Problems in the Biological Sciences*, Proceedings of Symposia in Applied Mathematics, 14, pp. 117-158, American Mathematical Society, Providence, R.I.

[6] Adekoya, A.F., Akinwale, A.T. and Oke, O.E., (2008). A medical expert system for managing tropical diseases. *Proceedings of the Third Conference on Science and National Development*, 74-86.

[7] Oluwagbemi, O., Adeoye, E., Fatumo, S., (2009). Building a computer-based expert system for malaria environmental diagnosis: An alternative malaria control Strategy, *Egyptian Computer Science Journal*, 33(2009), 55-68.

[8] Obot, O.U., and Uzoka, F.M.E., (2008). Fuzzy rule-based framework for the management of tropical diseases, *International Journal of Engineering and Informatics*, 1(2008), 7-17.

[9] Awotunde, J.B., Matiluko, O.E., & Fatai, O.W. (2014). Medical Diagnosis System Using Fuzzy Logic, African Journal of Computing and ICT, 7(2), 99-106, Published by IEEE Computer Society, Nigeria Section.

[10] Gilles MT, De Meillion (1968). The Anophelinae of Africa south of the Sahara. South African Institute for Medical Research, Publication 54.

[11] Detinova TS, (1962). Age-grouping methods in Diptera of medical importance, with special reference to some vectors of malaria. Geneva, World Health Organization, 1962 (WHO Monograph Series No. 47).

[12] Martens WJM (1997). Health impacts of climate change and ozone depletion: An eco-epidemiological modelling approach. PhD thesis, Maastrict University, Netherlands, 1997.

[13] Worrall E, Rietveld A, Delacollette C, (2004): The burden of malaria epidemics and cost-effectiveness of interventions in epidemic situations in Africa. American Journal of Tropical Medicine and Hygiene 2004, 71(Supplement 2):136-140.

[14] Khalid, .A. E., Elfatih .M. M., Tarig, A., Salah, .H. A., Abdalla, .H. E., Mahmoud, .A. M., (2009). Feasibility and acceptability of home-based management of malaria strategy adapted to Sudan's conditions using artemisinin-based combination therapy and rapid diagnostic test, Malaria Journal 2009, 8:39, doi:10.1186/1475-2875-8-39

[15] Utzinger, J., Yesim, T., Burton, S.H., (2001). Efficacy and Cost-effectiveness of Environmental Management for Malaria Control, *Tropical Medicine & International Health*, 6(2001), 677-687.

[16] E. S. Berner, R. S. Maisiak, C. G. Cobbs, and O. D. Taunton, (1999). Effects of a decision support system on physicians' diagnostic performance. Journal of the American Medical Informatics Association : JAMIA, 6(5):420–427.

[17] A. Hall and G. Walton (2004). Information overload within the health care system: a literature review. Health Info Libr J, 21(2):102–108.

[18] J. Wyatt (1991). Use and sources of medical knowledge. Lancet, 338(8779):1368–1373.

[19] G. A. Miller (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev, 63(2):81–97, 1956.

[20] R. Glassman, K. Leniek, and T. Haegerich (1998). Human working memory capacity is 7 2 in a radial maze with distracting interruption: possible implication for neural mechanisms of declarative and implicit long-term memory. Brain Research Bulletin, 47(3):249–256.

[21] Jon Brassey, Glyn Elwyn, Chris Price, and Paul Kinnersley. Just in time information for clinicians: a questionnaire evaluation of the attract project. BMJ, 322(7285):529–530

[22] Richard Smith (1996). What clinical information do doctors need? BMJ, 313(7064):1062–1068.

[23] Gordon D. Schiff, Seijeoung Kim, Richard Abrams, Karen Cosby, Bruce Lambert, Arthur S. Elstein, Scott Hasler, Nela Krosnjar, Richard Odwazny, Mary F. Wisniewski, and Robert A. McNutt (2005). Diagnosing Diagnosis Errors: Lessons from a Multi-institutional Collaborative Project. Advances in Patient Safety;2:255-278., vol 2, Pp 255–278.

[24] Mark L. Graber, Nancy Franklin, and Ruthanna Gordon (2005). Diagnostic error in internal medicine. Arch Intern Med, 165(13):1493–1499.

[25] Faith-Michael Emeka Uzoka, Okure Obot, Ken Barker, J. Osuji, (2011). An experimental comparison of fuzzy logic and analytic hierarchy process for medical decision support systems, computer methods and programs in biomedicine 103, 10-27.

[26] Obot, O.U., Uzoka, F.M.E., Barker, K. and Osuji, J., (2010). An experimental comparison of fuzzy logic and analytic hierarchy process for medical decision support systems. *Computer Methods and Programs in Biomedicine*, 6, (2010), 123-130.

[27] Uzoka F.M. and Barker K. (2005). Medical Decision Support System using Analytic Hierarchy Process: A case study of malaria diagnosis

[28] Jimoh R.G., Awotunde J.B., Babatunde A. O., Ameen, A. O , James T.R. & Fatai O.W. (2014). Simulation of Medical Diagnosis System for Malaria Using Fuzzy Logic, International Journal of Information Processing and Communication (IJIPC), Vol. 2 Issue (1), 59-68, published by Faculty of Communication and Information Sciences, University of Ilorin, Ilorin.

[29] Mir Anamul Hasan, Khaja Md. Sher-E-Alam and Ahsan Raja Chowdhury, (2010). Human Disease Diagnosis Using a Fuzzy Expert System, Journal of Computing, Vol 2, Pp. 66-70.

[30] Djam, X.Y., Wajiga, G. M., Kimbi Y. H. and Blamah, N.V., (2011). A Fuzzy Expert System for the Management of Malaria, International Journal of Pure and Applied Sciences and Technology ISSN 2229 – 6107, 5(2) (2011), pp. 84-108.

[31] Ojeme Blessing Onuwa (2014). Fuzzy Expert System for Malaria Diagnosis, Oriental Journal of Computer Science & Technology, Vol.7, No. (2):Pgs. 273-284.

[32] Nidhi Mishra and .P. Jha (2014). A Review on the Applications of Fuzzy Expert System for Disease Diagnosis, International Journal of Advanced Research in Engineering and Applied Sciences, Vol. 3, No. 12, Pp. 28-43.

[33] Akinyokun O. C., Iwasokun G. B., Arekete S. A, and Samuel R. W., (2015). Fuzzy logic-driven expert system for the diagnosis of heart failure disease, Artificial Intelligence Research, Vol. 4, No. 1, Pp. 12-21.

[34] O.W. Samuel, M.O. Omisore, B.A. Ojokoh, (2013). A web based decision support system driven by fuzzy logic for the diagnosis of typhoid fever, Expert Systems with Applications 40 (2013) 4164–4171.

[35] Priynka Sharma, Singh DBU, Manoj Kumar Bandil and Nidhi Mishra (2013), " Decision Support System for Malaria and Dengue Disease Diagnosis (ASSMD), International Journal of Information and Communication Technology, volume 3, Number 7, Pg 633-640.

[36] Olabiyisi S.O, Omidiora E.O,Olaniyan M.O and Derikoma (2011)," A Decision Support System Model for Diagnosing Tropical Diseases Using Fuzzy Logic", African Journal of Computing & ICT Vol4 No2 issue 2, Pg 1-6.

[37] Bruno B Andrade, Antonio Reis-Filho, Austeclino M Barros, Sebastião M Souza-Neto, Lucas L Nogueira, Kiyoshi F Fukutani, Erney P Camargo, Luís MA Camargo, Aldina Barral, Ângelo Duarte and Manoel Barral-Netto, (2010). Towards a precise test for malaria diagnosis in the Brazilian Amazon: comparison among field microscopy, a rapid diagnostic test, nested PCR, and a computational expert system based on artificial neural networks, *Malaria Journal* 2010, **9**:117

[38] Shruti Annaldas and Shirgan S.S (2015). Automatic Diagnosis of Malaria Parasites Using Neural Network and Support Vector Machine, International Journal of Advanced Foundation in Computer(IJAFRC), Volume 2, Special Issue (NCRTIT 2015), Pg 62 -65.

[39] David N. B., Robi N. M., Neil A. S., Wilbur A. L., and Daniel A. F., (2009). Mobile Phone Based Clinical Microscopy for Global Health Applications, Breslauer DN, Maamari RN, Switz NA, Lam WA, Fletcher DA (2009) Mobile Phone Based Clinical Microscopy for Global Health Applications. PLoS ONE 4(7): e6320.

[40] World Health Organization, (2010). Guidelines for the treatment of malaria, 2nd Edition,

[41] Oguntimilehin A, Adetunmbi A.O . Olatunji K.A., (2014). A Machine Learning Based Clinical Decision Support System for Diagnosis and Treatment of Typhoid Fever, International Journal of Advanced Research in Computer Science and Software Engineering, Vol4, I6 Pp. 961-969.

[42] Kamukama I (2009), "Clinical Protocol-Based Decision Support System for Malaria Treatment",M.Sc Computer Science Thesis, Makere University.

[43] Nicholas E. R., Charles J. P., David M. R., and Adriano G. D., (2006). Automated image processing method for the diagnosis and classification of malaria on thin blood smears, Medical and Biological Engineering and Computing, Vol 44, Issue 5, pp 427–436.

[44] Gloria Díaz, Fabio A. González, Eduardo Romero, (2010). A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images, Journal of Biomedical Informatics 42 Pp. 296–307.

[45] Suryadeep Roy, S.S Tripathy, (2016). A Novel Approach for the Detection of Malaria Parasites and Measure its Severity Using Image Processing and Fuzzy Logic, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 5, I4, Pp. 3081-3088.

[46] D. Anggraini, A. S. Nugroho, C. Pratama, I. E. Rozi, Aulia Arif Iskandar and Reggio Nurtanio Hartono, (2011). Automated status identification of microscopic images obtained from malaria thin blood smears, *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*, Bandung, 2011, pp. 1-6.

[47] J.E. Arco, J.M. Górriz, J. Ramírez, I. Álvarez, and C.G. Puntonet, (2015). Digital image analysis for automatic enumeration of malaria parasites using morphological operations, Expert Systems with Applications 42 Pp. 3041–3047.

[48] Ugwu C, N.L Onyejegbu and Obagbua I.C (2009), "The Application of Machine Learning Technique for Malaria Diagnosis", Nigeria Computer Society 23rd National Conference pg 151- 158

[49] Sneha Narayan Chavan & Ashok manikchand (2014), "Malaria Disease Identification and Analysis using Image Processing, " International Journal of Latest Trends in Engineering and Technology (IJLTET) Vol.3 issue 3.

[50] Vipul Parkhi, Pooja Pawar and Archana Surve (2013), " Computer Automation for Malaria Parasite Detection using Linear Programming", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering , Vol.2 issue 5,Pg 1984-1988.

[51] Steve, A., Joseph, W, Susan, J. H.,Gary, M. K. (2006). Electronic Health Records: *A Global Perspective Overview*

[52] Gunter, Tracy D; Terry, Nicolas P (2005). "The Emergence of National Electronic Health Record Architectures in the United States and Australia: Models, Costs, and Questions". *Journal of Medical Internet Research*

[53] Weddi, D. (2005). Using information systems to manage health in Uganda: *2004- 2005Annual Health Sector Performance report*.www.iconnectonline.org

[54] O'Brien J. A. (2002).Management Information Systems: Management information Technology in the e-business enterprise. Tata McGraw-Hill, New Delhi.

[55] CDC Treatment Guidelines, (2013). Treatment of Malaria (Guidelines for Clinicians), Malaria Case Surveillance Report July, 2013 Pp. 1-8.

[56] Noppadon Tangpukdee, Chatnapa Duangdee, Polrat Wilairatana and Srivicha Krudsood, (2009): Malaria Diagnosis: A Brief Review, *Korean J Parasitol. Vol. 47, No. 2: 93-102.*

[57] Bell DR, Jorgensen P, Christophel EM, Palmer KL (2005). Malaria risk: estimation of the malaria burden. Nature; 437: E3-E4.

[58] Reyburn H, Mbakilwa H, Mwangi R, Mwerinde O, Olomi R, Drakeley C, Whitty CJ, (2007). Rapid diagnostic tests compared with malaria microscopy for guiding outpatient treatment of febrile illness in Tanzania: randomised trial. BMJ; 334: 403.

[59] Looareesuwan S. Malaria. In: Looareesuwan S, Wilairatana Peds, (1999). Clinical Tropical Medicine. 1st ed. Bangkok, Thailand. Medical Media, p 5-10.

[60] Mwangi TW, Mohammed M, Dayo H, Snow RW, Marsh K, (2005). Clinical algorithms for malaria diagnosis lack utility among people of different age groups. Trop Med Int Health; 10: 530-536.

[61] McMorrow ML, Masanja MI, Abdulla SM, Kahigwa E, Kachur SP, (2008). Challenges in routine implementation and quality control of rapid diagnostic tests for malaria-Rufiji District, Tanzania. Am J Trop Med Hyg; 79: 385-390.

[62] T. O. Oladele, J. S. Sadiku & R. O. Oladele (2014). Coactive Neuro-Fuzzy Expert System: A Framework for Diagnosis of Malaria. Afr J. of Comp & ICTs. Vol 7, No. 2. Pp 173-188.

[63] A.T. Owoseni & I.O. Ogundahunsi, (2016). Mobile-Based Fuzzy Expert System for Diagnosing Malaria (MFES), I.J. *Information Engineering and Electronic Business*, 2, 14-22.

[64] A. B. Adehor, and P. R. Burrell, (2008). The Integrated management of health care strategies and differential diagnosis by expert system technology: A single-dimensional approach, *World Academy of Science, Engineering and Technology*, 2008, 44, pp. 533-538.

[65] M. E. Rafael, T. Rerrie, A. Magill, Y. Lim, F. Girosi, and R. Allan, "Reducing the burden of childhood malaria in Africa: the role of improved diagnostics", *Nature Publishing Group*, pp. 39-48.

[66] D. Chandramohan, I. Carneiro, A. Kavishwar, R. Brugha, V. Desai, and B. A. Greenwood, (2001). Clinical algorithm for the diagnosis of malaria: results of an evaluation in an area of low endemicity, *Tropical Medicine and International Health*, 6(7), pp. 505-510.

[67] K. A. Bojang, S. Obaro, L. A. Morison, and B. M. Greenwood, (2001). A prospective evaluation of a clinical algorithm for diagnosis of malaria in Gambian Children, *Tropical Medicine and International Health*, 5(4), pp. 231-236.

[68] A. O. Adetunmbi, A. Oguntimilehin, and S.O. Falaki, (2012). Web-Based Medical Assistant System for Malaria Diagnosis and Therapy, *GESJ: Computer Science and Telecommunications*, 1(33), pp. 42-53.

[69] Prabhu S, Muhammed W, Asma A and Kathy A. J-T. (2003), "Decision Support Systems to identify different species of malaria parasites

[70] Sio SW, Sun W, Kumar S, Bin WZ, Tan SS, Ong SH, Kikuchi H, Oshima Y, and Tan KS. (2007). MalariaCount: an image analysis-based program for the accurate determination of parasitemia, J Microbiol Methods. Vol 68, Isusse, pp. 111-8.

Medical Diagnosis System for Ebola Virus using Fuzzy Logic

Rasheed Gbenga JIMOH
Department of Computer Science,
University of Ilorin
Ilorin, Kwara State, Nigeria.
jimoh_rasheed@yahoo.com

Joseph Bamidele AWOTUNDE
Department of Computer Science,
University of Ilorin
Ilorin, Kwara State, Nigeria.
jabonnetbylinks@gmail.com

Adekunle Andrew AFOLAYAN
Department of Computer Science,
Kwara State College of Education,
Oro, Kwara State
andrewafolayan1@gmail.com

Opeyemi Emmanuel MATILUKO
Center for Systems and Information Services
Landmark University
Omu-Aran, Kwara State, Nigeria.
opmat01@yahoo.com, matiluko.opeyemi@lmu.edu.ng

Abstract: **Expert System as a branch of Artificial Intelligence is finding popularity and reputation in the field of medicine especially in medical support diagnosis system and across all others disciplinary. Many researchers have tried to addresses the problems of outdated method of medical diagnostic that full of inaccuracy and imprecision. Others researchers have made an attempt to design and implemented several algorithms using soft computing such as Fuzzy Logic, Genetic Algorithms and even Artificial Neural Network in address the problems of inaccuracy and imprecision in the field of medicine. Because of been afraid of direct contact with a patient with Ebola virus, a medical doctor is always being careful in handling cases of Ebola diseases. This paper present a rule based diagnosis system for Ebola virus using fuzzy logic. The medical expert system takes user input and depending on the input (symptoms) of the patient, diagnoses if the patient is suffering from Ebola virus or not. The System categorizes the Ebola symptoms using four linguistic variables based on the existing symptoms. Fuzzy rules are expressed and applied by java programming language model for the Ebola diagnosis system. The system provides useful evidence and anticipated data to develop fuzzy logic based control system for enhancement of the diagnosis of Ebola virus in real time application. The proposed system helps in the area of keeping patients medical records as well asking the Ebola patient to go for medical checkup for proper medication. Further work is expected to design, develop and implement the use of factor analysis and data mining techniques for classification and evaluation of qualified importance of the symptoms used in the process of Ebola diagnosis.**

Keyword: *Ebola virus, Algorithms, Diagnosis, Fuzzy logic, Medical.*

1.0 INTRODUCTION

Patients visiting hospital in other to get quickly, precise and adequate treatment and care without queuing or postponement of treatment from the best hands [1]. No patient wants to be used as a specimen from the hands of a doctor particularly those that are just finishing from medical school and have not gain or improved on the knowledge since graduation from school and patient do not want any delay in getting treatment for whatever reason [1]. The above-mentioned problems have been the major issues in most hospitals, especially in developing countries [2].

Many people didn't like visit the hospital because of the smell of drugs, hostile nurses, and long queue await them before they were allow to see doctor or simply laziness among others factors [3]. Patients are death in most developing countries just because of minor ailments that could be been treated easily [1] [4].

Another major problems is the scarce introduction and the infusion of modern technologies that can help the activities in the practice of modern medicine. Technology have been proved to be very useful tool in all areas of human endeavours that can helps in the areas of increase in efficiency, effectiveness and in turn productivity, the systematizing some of the medical processes involved in attending to patients that will increased the numbers of patients that can be attended to at the same time from different locations would be of enormous assistance [1]; [5].

Automated system are of greatly used in many developed world to improve the quality of medical services, it also reduced the workload of the medical doctors [6]; [7]; [8]. Artificial Intelligence (AI) as branch of computer science have been used to creating machines that can involve on characters that human consider special or intelligent [9]. The competence to create intelligent machines has interested humans since

ancient times and with the new interest in the creation of smart machines coping with the research into AI programming techniques, the dream of researchers are becoming a reality [10]; [8].

Researchers are creating systems which can caricaturist human thinking, recognize speech win the best chess player ever known ad many other works or things never possible before for a smart machine to do. There are intelligent machine now that can drive car without any human supervision. It is of great important to find out how the military is applying AI logic to its works and hi-tech systems for you to be able to appreciate the work of these smart machine, and to see in the near future how the world of AI may impact our lives and human intelligent [8]; [11]. The used of human expert and over dependence on human expertise can be reduced if his/her skill/knowledge can be shared/shifted into a computer system [12]. The work proposed a supporting system that will deal with the problems of Ebola diagnosis using an (ES). ES make use of human knowledge and reproduced it into computer to solve problems that customarily entail human expertise [13]; [14]. The creation of ES pursues and exploits appropriate and useful information from their human users by make used of available knowledge bases in the system to make useful recommendation [15]. The user can store heuristic knowledge using ES. Java programming language was used in development of the expert system [16]; [17]. Java programming was also used to design the system because of it flexibility in usage and to simplify the improvement of software to model human knowledge or expertise for medical therapy.

The human health gives us the opportunity and sense of self-reliance, allowing us to adventure more about the surrounding world that we would have done without good health. As it is commonly said that a healthy nation is a wealthy nation, one will be more productive both in reasons and otherwise when he/she is sound in health. Without good health no organ of the body will be function perfectly well.

Ebola has been proved to be terminal diseases that many people fear and very careful to have direct contact with the infected person, because having direct contact can easily transfer the disease to the contacted person. This paper addresses the problems of diagnosing Ebola disease and shortening the period of detection and incubation of Ebola virus.

1.1 Ebola Virus

The Ebola virus previously called Ebola hemorrhagic fever was discovered in rural Democratic Republic of Congo (formerly called Zaire) in the year 1976 [18][19][20]. Then Ebola virus had mostly be restrained to relatively small outbreaks in the rural area, until in the year 2014 that the outbreak hit the urban part of Liberia, Sierra Leone, and Guinea [20]. The world health Organization (WHO) has been condemned because of its low level of commitment in response to the epidemic in the affected areas has called this the most severe critical health disaster seen in modern times [18]. It is very difficult to differentiate between EVD from other infectious Filovirus hemorrhagic fever such as laser fever, hemorrhagic fever, typhoid fever and meningitis, which some of them have the same symptoms/signs. Collecting samples from Ebola victims are very extreme biohazard risk, but the laboratory testing on the unconfirmed and non-inactivated samples should be conducted under extremely cares and in a biological containment conditions [20][19].

The Ebola virus causes serious illness and severe problems that is often fatal if not contained on time or leave on untreated. Ebola virus disease (EVD) was first discovers in 1976 in two (2) instantaneous epidemics, one in Sudan in a village called Nzara, and in another village called Yambuku, in Democratic Republic of Congo. The EVD later arisen in a village near the Ebola river where the ailment take its name from [21][20].

The most recent and current epidemic of EVD in West Africa, was occurred in March 2014, which recorded to be the largest and most intricate Ebola outbreak since the discovery of Ebola virus in 1976 [19]. The outbreak recorded more cases and deaths than all others outbreak combined together. The outbreak spread within the countries because of movement of the carriers of this virus from one country to another starting in Guinea and the virus spreading across countries starting from Sierra Leone and Liberia then travel by air to Nigeria (one traveler), and to Senegal by land (one traveler) [21].

The virus called Ebola can be contained, controlled or limited to an area by imposing or enforcing quarantines the carriers by separating the sick patient from the healthy society, and keeping the exposed people away from the general population for the virus's twenty-one day incubation period for proper observation of the carrier of the virus. Every traveler especially from the affected countries are potential Ebola carriers that can be screened for symptoms and travel history, which is the currently case for travelers from West Africa to the United States. This must be done regularly so as to contain the virus from spreading to other part of the world, which can lead to serious health problems for the entire world. Healthcare workers must always use protective equipment mainly created and designed to prevent, contain or resist the

virus. The United State Center for Disease Control has come up with a revised guideline, which has be released in October to prevent misuse of the kit/equipment, this call for regulations training and supervision on the doming and stripping of the outfits [21].

Other treatment for Ebola virus and measures have proven not reliable and more controversial, their efficacy is further questionable [22]. The affected countries of the world are Guinea, Sierra Leone and Liberia all the countries are from Africa continent and have very weak health systems, deficient in human and infrastructural resources to fight the epidemic, having only recently emerged from long periods of conflict and instability [23]. "On August 8, the WHO Director-General declared this outbreak a Public Health Emergency of international Concern" [19].

The time interval from infection with the virus of the patient to beginning of symptoms been manifested is from 2 to 21 days, which is the incubation period. A patient is not declared infections until they develop symptoms. The most important and first symptoms are the sudden onset of fever fatigue, muscle pain, headache and sore throat [21]. After which the following symptoms will be manifested vomiting, diarrhea, rash, symptoms of impaired kidney and liver function, and the most deadly case which occurred lastly in some cases, are both internal and external bleeding (e.g. oozing from the gums, blood in the stools) [18][20][19][22]. The laboratory findings of Ebola virus include low white blood cell and platelet counts and elevated liver enzymes [23].

1.2 Fuzzy Logic

The first recorded used of fuzzy logic began with the 1965 proposal of fuzzy set theory by Lotfi Zadeh. [24][25] fuzzy logic has been well used and applied in many fields, from it's used in control theory to the used in AI [24]. Both fuzzy logic and probabilistic logic are mathematically very similar because both have truth values that alternating between 0 and 1 but theoretically different, due to different in their interpretations, the interpretations of probability theory in fuzzy logic corresponds/resembles to "degrees of truth", while probabilistic logic yield different types of models of the same real-world situations/conditions [26].

It is a common saying and widely accepted that the main braches and components of soft computing are fuzzy logic, probabilistic reasoning, neural computing, and genetic algorithms [27], all these are also braches of AI algorithms [24]. The used of fuzzy logic was adopted to implement the proposed system since it is a powerful tool to deal with the problem of uncertainty and imprecision that always occurred in the field of medicine sciences. It is powerful tools that can be used in an uncertain cost-effective decision setting in deal with the imprecision and ambiguity of human thought and the difficulties in approximating inputs [28][29][25]. It has been used to bridge the gap between traditional and outdated approaches of diagnosis and computer-assisted diagnosis by handling the issues of vagueness, imprecision and ambiguity inherent[27][28] in the field of medicine which is always called an expert system medical diagnosis or medical decision support system. The system are every trustworthy amidst medical doctors know that their medical knowledge and resulting diagnosis are not always correct and is full of uncertainty with imprecise formulations[27][28].

2.0 OVERVIEW OF THE METHOD

The data for this research will be collected using patients data collected from the hospital, the methods of data collection involves direct interview of experts and consultants in general medicine. The useful data for the research that will be collected include various signs and symptoms of the virus and various ways using in diagnosing Ebola virus. Also, several journal articles, paper presentations and relevant past research and projects were consulted in order to have a direction for this research, all these will be used stimulate the system using Java Programming language. The UITH, Ilorin (University of Ilorin Teaching Hospital, Ilorin, Kwara State) will be used for data collection.

3.0 SYSTEM DESIGN AND FUZZY LOGIC STRUCTURE

Over the years model, techniques and approaches for developing a software has emerged, these includes the use of CASE based tools or software. This paper is aimed at designing an ES system for the diagnosis of Ebola using the system follow the procedure for a design of a standard rule based expert system also using fuzzy logic. A fuzzy inference method was used to deduce the percentage risk of the patient.

Logical (iSkinEruptionlementation independent): These aspects of the system that can be designed without knowledge of the iSkinEruptionlementation platform.

Physical (iSkinEruptionlementation dependent): It is the other aspects of the system that are very dependent on the iSkinEruptionlementation portion and platform that will be used.

After the review of the existing system and identifying its shortcomings, it has been concluded that there is a need to automate the system by creating software that will perform the various operations done by the existing system in a more precise and efficient way.

3.1 History Taking

Significantly defining the reason for a patient's consultation is required. In this stage the interpersonal services of the medical staff plays a big role in facilitating communication between the two parties. At this early stage a good and responsible medical doctor starting with an open question like "How can I help you today" and it of important here to pay an attention to non-verbal cues that are more likely to be productive than launching into a closed question and answer session.

3.2 Basic Data

Patient's Name: This identifies the patient in context for the purpose of interaction.

Age of patient: Some diseases are always peculiar to a particular age range. The age of the patient will help the doctor to know what kinds of questions to ask based on the age range.

Sex: There are some diseases that span across both sexes. The sex of a patient also goes a long way in determining the types of questions to ask the patient.

Occupation: The occupation of the patient will also be recorded.

3.3 Awkward CoSkinEruptionlaints

During the process of CoSkinEruptionlaints it is necessary to ensure that the patient does not feel uncomfortable as not to be able to give all the necessary information needed for accurate diagnosis. There are two major things to put in consideration here:

Firstly, the types of questions to asked the patients and how to put it across so that the patients will not feel un-protective around you. Secondly always make the patients comfortable around by asking open questions that need direct answers not a closed question, which can make the patient not response appropriately.

This process is highly disturbing by its very nature. The patient has been exposed through the questions you have asked, which maybe both literally and figuratively of the layers that protect them from the physical and psychological probes of the outside world hence the patients might feel a sense of insecurity and might not be able to answer you with easy mind.

The patient must be asked exhaustive, cherished questions. At this stage of coSkinEruptionletely that is at odds with the normal day-to-day interaction. There is no way the medical doctor will proceed without asking questions, the doctor have to scrutinizing into the life of the patient else coSkinEruptionletely become stranger. But it can be done in a way that the questions asked are in line with the problems at hands and should always maintains respect for the patient's dignity and privacy. Many doctors/nurses develop resistant to the sense that they are violating a patient's personal space and can unconsciously and unintentionally over step boundaries by the type of questions asked during consultation. Sidestepping by avoiding this types of questions is not an easy task. Always Listening and responding appropriately to the internal warnings within you can help to physique the normal interactions will help overcome this barrier.

3.4 Fuzzification Process

Fuzzification process is the fluctuating and changing of a real scalar value into fuzzy value [30][31][32]. The process can be achieved by fuzzifiers, its generally categorized into four types namely: Traingular fuzzifiers, Singleton fuzzifiers, Trapezoidal fuzzifiers and Gaussian fuzzifier [33][34]. The Triangular fuzzifier will be used in the research.

Fuzzifier of data using Triangular Fuzzifier is carried out by selecting input parameters, and sticking them to the upper boundary of relationship function to determine the degree of membership. The first step in design and development of fuzzy logic based system is to construct fuzzy sets for the parameters to be used in the expert system. The development of fuzzy logic in constructing fuzzy sets parameters are shown in equations (1) to (4). The input and output parameters that were used and selected for this research were describe with the four linguistic four adaptable variables. The numeric range values of fuzzy sets for each linguistic is shown in table 1:

Table 1: Range of fuzzy Values

Linguistic Variables	Fuzzy Values
Mild	$0.1 \leq x \leq 0.3$
Moderate	$0.3 \leq x \leq 0.6$
Severe	$0.6 \leq x \leq 0.8$
Very Severe	$0.8 \leq x \leq 1.0$

Fuzzification starts with the conversion of the raw data from its present state. During the process, the Linguistic variables used are evaluated using the selected fuzzifier membership functions and this are the associated variables that are accoSkinEruptionanied by degree of membership ranging from 0 to 1 as shown in equation (1) to (4). The formulas that were used were formulated by help of both the expert doctor in the field of tropical medicine and literature.

$$\mu mild(x) = \begin{cases} 0 \ if \ x \ \leq 0.1 \\ \frac{x-0.1}{0.2} \ if \ 0.1 \ \leq x \ \leq 0.3 \\ \frac{0.2-x}{0.2} \ if \ 0.2 \ \leq x \ \leq 0.3 \\ 0 \ if \ x \ \geq 0.2 \end{cases} \quad \text{Eq. 1}$$

$$\mu moderate(x) = \begin{cases} 0 \ if \ x \ \leq 0.3 \\ \frac{x-0.3}{0.2} \ if \ 0.3 \ \leq x \ \leq 0.6 \\ \frac{0.45-x}{0.2} \ if \ 0.45 \ \leq x \ \leq 0.6 \\ 0 \ if \ x \ \geq 0.45 \end{cases} \quad \text{Eq. 2}$$

$$\mu severe(x) = \begin{cases} 0 \ if \ x \ \leq 0.5 \\ \frac{x-0.45}{0.2} \ if \ 0.6 \ \leq x \ \leq 0.8 \\ \frac{0.45-x}{0.2} \ if \ 0.7 \ \leq x \ \leq 0.8 \\ 0 \ if \ x \ \geq 0.7 \end{cases} \quad \text{Eq. 3}$$

$$\mu very\ severe(x) = \begin{cases} 0 \ if \ x \ \leq 0.8 \\ \frac{x-0.1}{0.2} \ if \ 0.8 \ \leq x \ \leq 1.0 \\ \frac{0.2-x}{0.1} \ if \ 0.9 \ \leq x \ \leq 1.0 \\ 0 \ if \ x \ \geq 1.0 \end{cases} \quad \text{Eq. 4}$$

After the formulation of the equations that will be used, the next step in the process is the development of fuzzy rules. The fuzzy rules were designed with the support of five medical doctors who are experts in the selected field. The knowledge-based of FESMM was designed with the aid of combination theory that has no many fuzzy rules: only the valid rules where selected by the medical doctors.

An expert rule is said to fire if any of the antecedence parameters, which are mild, moderate, severe, and very severe evaluate to true (1), other if all the parameters evaluate to false (0), it does not fire.

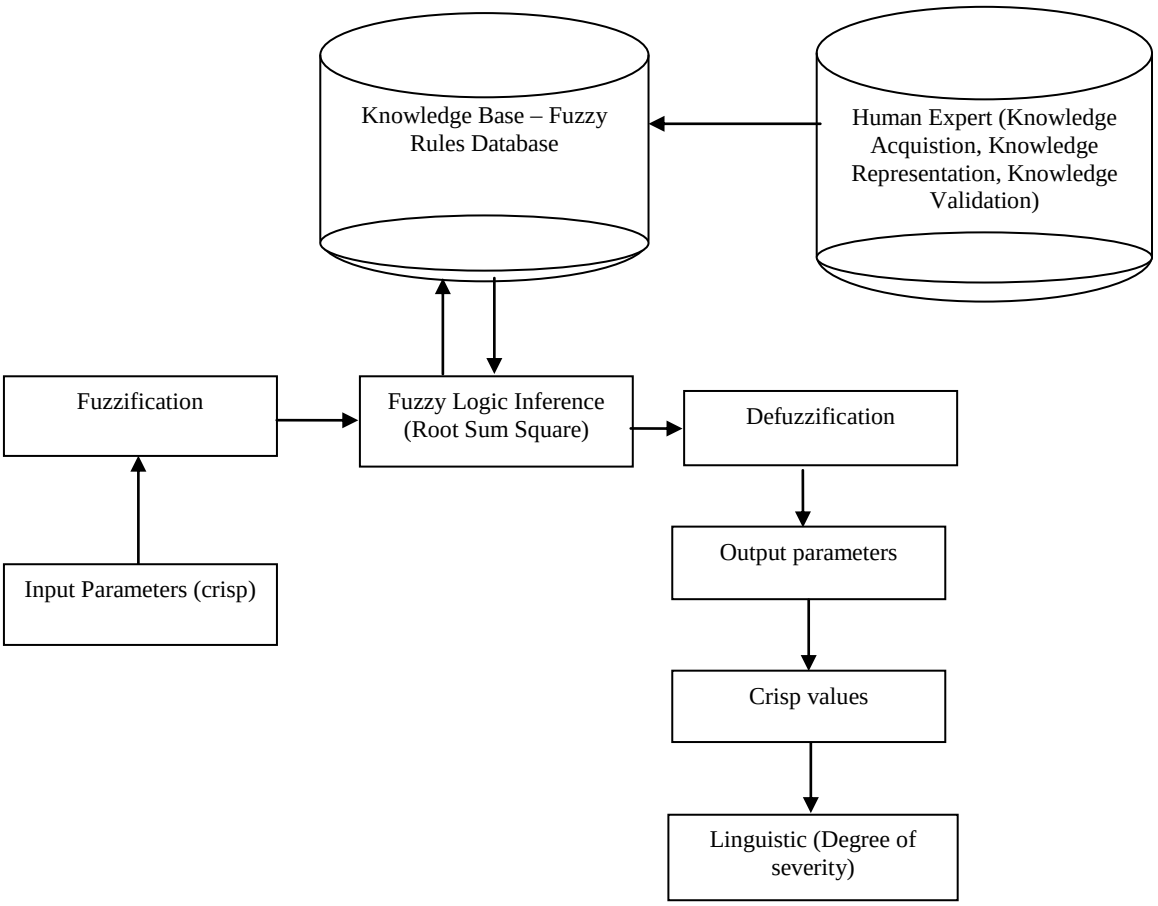


Figure 3.1: Fuzzification process

4.0 FUZZY INFERENCE

Inference is the process of drawing conclusion from existing data. Fuzzy inference is the procedure of drawing conclusion from a given input to an output using the theory of fuzzy sets rules in the knowledge based system. The fundamental of decision-making from the input to output is process in which the inference engine make used of the rules formulated and contained in the rule base. The fuzzy inference engine draws conclusion base on the rules in the knowledge base system, the final decision is derives base on the rules. The FESMM inference engine will make uses a forward chaining machine to examine the knowledge for the sySkinEruptions of a disease. For each, the inference mechanism looks up the membership values based on the condition of the rule. The fuzzy inputs are normally mapped into their corresponding weighting factors and their related linguistic variables to determine their degree of association, membership and their relationship. Aggregation operator in fuzzification process is sued to calculate the degree of fulfillment or firing strength of a rule.

The research paper decided to apply fuzzy logical AND gate evaluate the coSkinEruptionosite firing strength of the rules in the knowledge base system. In practice, the fuzzy rules sets always have several antecedents sets that are combined using fuzzy logical operators which is alos known as logic gate such as AND, OR, and NOT, though their definitions tend to vary AND siSkinEruptionly uses minimum weight of all the antecedents, while the uses of OR have maximum value. "There is also the NOT operator that subtracts a membership function from 1 to give the THEN part is called the consequent".

The authors will make use of the AND operator to combine the antecedent parts of the rules. The membership of truth (R) of the rules are determined for each rule by evaluating the non-zero minimum values using the AND operator. The inference engine evaluate all the rules in the rules sets by combines the weighted consequences of all the applicable (fired) into a single fuzzy set. The inference engine technique used in the work is eSkinEruptionloyed that is the Root Sum Square (RSS). RSS is given by the formula in equation (5).

$$\sqrt{\sum R^2} = \left(R_1^2 \sqrt{R_2^2 + R_3^2 + R_4^2 + KR_n^2}\right) \qquad \text{eq. 5}$$

Where $R_1^2 + R_2^2 + R_3^2 + KR_n^2$ are truth strength values of the different rules used, which share the same inference and conclusion i.e. R = value of firing rule. RSS used the formulated rules and combines the effects of all applicable rules, scales the functions at their respective magnitudes for the parameters and for coSkinEruptionute which is the "fuzzy" centroid of the coSkinEruptionosate area. The method is more coSkinEruptionlicated mathematically than other formally used methods, this methods was selected for this paper since it gives the best-weighted influence in firing all rules. The methods RRS for drawing inference was found to be the most applicable and suitable method to infer data from the rules developed.

4.1 DEFUZZIFICATION

The main function of defuzzification process is to translate the output from the inference engine into crisp output [27][28]. The output from the inference engine is usually the fuzzy set and for almost all medical applications crisp values are essential. There are many types of the defuzzification methods namely: center-of-area (gravity), center-of-sums, mean of maxima and max-critenon. Abraham and Nath [35] pointed out that the Max-critenon produces the point by which the likelihood distribution of the action reaches a maximum value at the point of siSkinEruptionlest to iSkinEruptionlement. Most authors normally used center-of-area (center-of-gravity) or centroid method because when it is used, the defuzzified values tend to move efficiently and smoothly around the fuzzy output region [36][37]. The two techniques always gives more accurate representation of fuzzy set of any shape [38][39]. The center-of-gravity (CoG) is used in a situation the fuzzy sets is discrete variables, the CoG Y can be estimated to overtake its shortcoming, and this is represented in equation 6 [27]. The weighted average of the centers of the fuzzy set was formulated and used instead of integration [28]. The CoG always used for the averaging technique. The CoG formula for defuzzification is given below:

$$\text{CoG } (Y') = \frac{\sum \mu_Y(x_i) x_i}{\sum \mu_Y(x_i)} \qquad \text{Eq. 6}$$

Where $\mu_Y(x)$= membership value function and X_i center of function.

The technique was used in this research paper since it is coSkinEruptionutationally siSkinEruptionle and intuitively conceivable.

5.0 DISCUSSION

Fuzzy logic is used to deal with the problems of uncertainty and vague terms, the methods of soft computing is widely accepted in different field of studies and sphere of life especially in the field medicine. The paper presented a novel method for diagnosis of Ebola virus in any patient that having

the symptoms of this deadly disease. Based on the selected area, the symptoms used will be select by the patient from the system, the Ebola virus symptoms were stored in the system, the user needs to select symptoms based on how the patient are feel. The next step for the system after the selection of symptoms, the patient will be asked some other related and vital questions based on the selected symptoms. The user must answers all the questions that the system asked, and the system response based on the answer selected. The fuzzy diagnosis system will diagnosis diseases and response based on its knowledge that is stored in the knowledge-base, it very necessary to add catalyst factor (if any), and to do the ranking before gives the result in fuzzy form. The available symptoms on the system enables the users to determine his/her possible ailments very quickly with the fuzzy sets on the system. The system was created based on the feedback from expert and specialist doctors, patients can also trust the result of the proposed system.

The purpose of the paper is to stimulate the concept of fuzzy rule based system that incorporated fuzzy techniques in simplifying medical diagnosis system using Ebola virus as a case study. The fuzzy expert system for diagnosis Ebola virus was designed. Following the fuzzy logic implementation, the paper make used of the selection of fuzzifier, inference engine and rule base techniques to determine the output of the fuzzy logic system. Triangular fuzzifier was used, the expert doctors was very helpful in the designing of the rule base system, and RSS inference method, and the authors engaged the fuzzy logic overcome the traditional ways of diagnosis Ebola virus and also to remove uncertainty, ambiguity and vagueness that characterized the oldest methods that was inherent in medical diagnosis.

Furthermore, the assign linguistics variables were used to diagnosis the degree of mildness, intensity or severity of the patient. The important of severity is to allow the medical doctors to assign diverse treatment to each of the patients according to their degree of analysis based on the results from the expert system. Among the soft computing techniques the fuzzy diagnosis has an advantage over other because it resembles human reasoning and human decision-making by its ability to work from rough reasoning and ultimately find a precise solution.

6.0 CONCLUSION

On the basis of all presented, it can be concluded that there is no doubt whether Fuzzy Expert Systems should be applied for medical purpose. The use of fuzzy logic have provides a proficient in dealing with the problems of inaccuracy and imprecision in medicine. The system is also a proficient way of arrived in reliable and precise diagnosis of Ebola more quickly and efficiently.

The simulation of fuzzy logic for Ebola virus medical diagnosis system provides quick and precise platform to help and assist in Ebola virus research, medical doctors and other health workers in Ebola endemic regions. Application of the system will go really assist in proper and efficiently help for diagnosing Ebola virus, and if used intelligently, the results will be very useful in any environment for diagnosing Ebola. Also the implementation of fuzzy logic based diagnostic system will reduced the risk of having direct contact with patient that have symptoms related to Ebola virus. This will also reduce the workload of medical doctors and ease other difficulties that are always faced during hospital consultation. The system should be imperiled to severities clinical examination before acceptability for the full development for uses as Ebola Diagnosis system.

7.0 FURTHER WORK

The application of factor analysis in the classification and evaluation of relative importance of the symptoms in the process of Ebola virus should be considered for further work. A clinical examination and evaluation of the system is very important before full acceptability in diagnosing Ebola virus.

REFERENCE

[1] S. a Fatumo, E. Adetiba, and J. O. Onaolapo, “Implementation of XpertMalTyph : An Expert System for Medical Diagnosis of the Complications of Malaria and Typhoid,” vol. 8, no. 5, pp. 34–40, 2013.

[2] P. D. Cleary, S. Edgman-Levitan, M. Roberts, T. W. Moloney, W. McMullen, J. D. Walker, and T. L. Delbanco, “Patients evaluate their hospital care: A national survey,” *Health Aff.*, vol. 10, no. 4, pp. 254–267, 1991.

[3] G. Powell-Cope, A. L. Nelson, and E. S. Patterson, *Patient Care Technology and Safety*. Agency for Healthcare Research and Quality (US), 2008.

[4] A. Coulter and P. D. Cleary, “Patients’ experiences with hospital care in five countries,” *Health Aff.*, vol. 20, no. 3, pp. 244–252, 2001.

[5] L. J. Frost and M. R. Reich, “Creating access to health technologies in poor countries,” *Health Affairs*, vol. 28, no. 4. pp. 962–973, 2009.

[6] E. Ortiz, C. M. Clancy, and AHRQ, “Use of information technology to improve the quality of health care in the United States.,” *Health Serv. Res.*, vol. 38, no. 2, pp. xi–xxii, Apr. 2003.

[7] K. K. Shaly and T. Shahana, “An Expert Selection System for Medical Diagnosis,” vol. 5, no. 6, pp. 2015–2017, 2016.

[8] S. S. A. Naser, A. Z. a Ola, and I. Technology, “an Expert System for Diagnosing Eye Diseases,” *Expert Syst.*, pp. 923–930, 2008.

[9] H. A. Sabri, “Making sense of knowledge management (KM), information technology (IT) and artificial intelligence (AI): an integrative approach,” *Asian Trans. Basic Appl. Sci.*, vol. 1, no. 4, pp. 1–16, 2011.

[10] O. Pimple, U. Saravane, and N. Gavankar, “Cognitive Learning Using Distributed Artificial Intelligence,” *Int. J. Mach. Learn. Comput.*, vol. 5, no. 1, pp. 7–11, Feb. 2015.

[11] J. R. Stuart and N. Peter, *Artificial Intelligence a Modern Approach*, vol. 72, no. 1–2. 2003.

[12] N. Asabere, "Future Work Can be enhanced by generating more possible disease at the same time each generation is done.Any updating o disease and treatment database can be done by installing new version of MDES system that replace the old version of System Medicine," *Int. J. Sci. …*, vol. 2, no. 6, pp. 333–344, 2012.
[13] X. Pan, "An experimental adaptive expert system," *Proc. Intersymp-2000'baden-bad.*, pp. 1–17, 2000.
[14] O. Azaab, S; Abu Naser, S; Sulisel, "A proposed expert system for selecting exploratory factor analysis procedures," *J. Coll. Educ.*, vol. 4, no. 2, pp. 9–26, 2000.
[15] L. E. F. Efraim Turban, *Expert Systems and Applied Artificial Intelligence*. Macmillan Pub. Co., 1992.
[16] R. Ahmad, "Expert Systems: Principles and Programming," *Scalable Computing: Practice and Experience*. PWS Pub. Co, p. 602, 2001.
[17] Jackson and P., "Introduction to expert systems," 1986.
[18] "WHO | Ebola virus disease," *WHO*, 2016.
[19] D. Renwick, "CFR Backgrounders," *U.S.-Cuba Relations*, pp. 1–5, 2016.
[20] H. Feldmann and T. W. Geisbert, "Ebola haemorrhagic fever.," *Lancet (London, England)*, vol. 377, no. 9768, pp. 849–62, Mar. 2011.
[21] E. M. Vingolo, G. A. Messano, S. Fragiotta, L. Spadea, and S. Petti, "Ocular Manifestations of Ebola Virus Disease: An Ophthalmologist's Guide to Prevent Infection and Panic," *BioMed Research International*, vol. 2015. Hindawi Publishing Corporation, p. 487073, 2015.
[22] A. D. Renwick and C. Editor, "Backgrounders Ebola Virus," pp. 1–5, 2014.
[23] L. Broxmeyer, "Is the Ebola virus real ?," vol. 2, no. 3, pp. 42–52, 2014.
[24] B. C. Arabacioglu and B. Cem, "Using fuzzy inference system for architectural space analysis," *Appl. Soft Comput.*, vol. 10, no. 3, pp. 926–937, Jun. 2010.
[25] K. Dutta and S. Sarthak, "Architectural space planning using evolutionary computing approaches: A review," *Artif. Intell. Rev.*, vol. 36, no. 4, pp. 311–321, 2011.
[26] L. Biacino and G. Gerla, "Closure Operators in Fuzzy Set Theory," Springer US, 1999, pp. 243–278.
[27] S. A. Oke, A. O. Johnson, I. O. Popoola, O. E. Charles-Owaba, and F. A. Oyawale, "Application of Fuzzy Logic Concept to Profitability Quantification in Plastic Recycling.pdf," *Pacific J. Sci. Technol.*, vol. 7, no. 2, pp. 163–175, 2006.
[28] A. O. Egwali, D. Ph, and J. C. O. M. Sc, "An Adaptive Neuro-Fuzzy Inference System for Diagnosis of EHF .," vol. 16, no. 1, pp. 251–261, 2015.
[29] R. G. Jimoh, J. B. Awotunde, O. W. Fatia, K. State, T. R. James, and E. State, "Simulation of Medical Diagnosis System for Malaria," vol. 5, no. 2, pp. 59–68, 2014.
[30] P. Singhala, D. N. Shah, and B. Patel, "Temperature Control using Fuzzy Logic," *Int. J. Instrum. Control Syst.*, vol. 4, no. 1, pp. 1–10, 2014.
[31] P. Kayal and S. Kannan, "Building Fuzzy Associative Classifier Using Fuzzy Values," vol. 3, no. 7, pp. 2012–2014, 2014.
[32] S. Princy, "Comparison of Triangular and Trapezoidal Fuzzy Membership Function," no. 8, pp. 46–51, 2016.
[33] P. Srinivas, "Comparative Analysis of Conventional Pid Controller and Fuzzy Controller with various Defuzzification Methods in a Three Tank Level Control System," *Int. J. Inf. Technol. Control Autom.*, vol. 2, no. 4, pp. 75–86, 2012.
[34] D. E. P. T. Of, M. E. E. Ngineering, F. E. U. Niversity, O. F. A. Griculture, A. Beokuta, D. I. O. F. W. Orks, S. Ervices, F. E. C. Ollege, O. F. E. Ducation, O. Siele, and A. Beokuta, "Modelling room cooling capacity with fuzzy logic procedure," vol. 35, no. 3, pp. 519–526, 2016.
[35] A. Abraham and B. Nath, "Hybrid intelligent systems design: A review of a decade of research," *… Trans-actions Syst. Man Cybern. (…*, 2000.
[36] D. Akay, M. A. Akcayol, and M. Kurt, "NEFCLASS based extraction of fuzzy rules and classification of risks of low back disorders," *Expert Syst. Appl.*, vol. 35, no. 4, pp. 2107–2112, Nov. 2008.
[37] S. Gollakota and D. Katabi, "Zigzag Decoding: Combating Hidden Terminals in Wireless Networks," *Acm Sigcomm*, vol. 38, no. 4, p. 159, 2008.
[38] L. H. Tsoukalas and R. E. Uhrig, *Fuzzy and neural approaches in engineering*. Wiley, 1997.
[39] B. Milenova, "Fuzzy and neural approaches in engineering," *Neural Networks*, vol. 10, no. 9, pp. 1740–1741, 1997.

Development of a Modified SIRA Model for Computer Virus Propagation

(SLIRA MODEL OF COMPUTER VIRUS SPREAD)

Peter M. Achuba*
Department of Mathematical Sciences,
Kogi State University (KSU)
Kogi State, Nigeria
*Email: peternuim85@gmail.com

Bamidele Oluwade**
Department of Mathematical Sciences,
Kogi State University (KSU)
Kogi State, Nigeria
**Email:deleoluwade@yahoo.com

***Abstract*: Computer viruses are capable of unleashing their effects (payloads) in two ways: either by means of the system users' action (e.g. Trojan horses) or by means of the encoded destructive sequence which enable it to infect systems without the aid of a system user (a typical example is a computer viral worm). The implication of the above is that computer systems can be latent infected before they become infectious. In this paper, the aim is to determine the behavior of computer viruses in a computer network which has fully functioning computer antivirus software installed on it. The model proposed is a modification of the SIRA (Susceptible, Infected, Recovered and Antidotal) model to include latent infected computers in the network which leads to the SLIRA (Susceptible, Latent infected, Infected, Recovered and Antidotal) model. In the model, some particular behaviors which some computer viruses exhibit (which lead to latent infection of computers by viruses) are considered. Thereafter, the disease-free, endemic equilibrium points, basal reproduction rate and stability conditions are sought. It can be inferred that every networked organization should have computer antivirus software installed on its system since whenever the total population of computers on the network is antidotal we achieve asymptotic stability.**

Keywords—Computer virus,SIS, SIR, SIRA model, SLIRA model, Antivirus software, Basal Reproduction Rate and First order ODE

V. INTRODUCTION

Computer viruses have been much of a bane to several individuals and organizations causing un-repairable damages and sometimes ones that cost lots of finance to bring about repair, other times, quite often, the issue of stolen valuable time, where an individual might spend several minutes waiting for a computer system to boot or for a particular program to run. Also, are situations of the data diddling virus where a thousand infected program modifies one bit in one randomly selected data file once a week. This implies that a thousand randomly selected bits of data in the infected system is being changed every week. A more humorous situation is one in which the virus creates typing errors whenever the user types faster than 60 words per minute.[1]

Also worthy to note, is the fact that epidemiology has being widely applied to the spread of computer virus in a computer network. Developing a mathematical model for the computer viral propagation is of critical importance not only for understanding better the behavior of computer viruses but also for stopping the spread of the virus. Due to the high similarity between computer virus and biological virus, some models for the spread of computer virus have been proposed. Say, [2], [3].

A reliable model of virus propagation is beneficial in the sense that, it allows researchers to better understand the threat posed by new attack vector and probably new propagation techniques. It would be beneficial to note that the use of conceptual models of worm propagation allowed researchers to predict the behavior of future malware, and later to verify that their predictions were correct. [4], [5]

It is a common phenomenon that when several individuals talk about the computer virus they actually refer to a particular type of scripting program but in true since the word "virus" mainly refers to viruses, worms and Trojan horse. Each of these individuals has profound characteristics which distinguishes them from the others i.e. the worm virus does not require the aid of a user to carry out its payload. This implies that the worm virus would spread faster than the normal computer viruses. And since the normal virus and Trojan horse virus actually needs the help of the machine user to activate its destructive sequence it would also mean that they would remain in a latent state unless aided by the user.

An issue with the SIS and SIR model would be that they lacked expression for computer systems which are latent infected and one which is in its antidotal state. Therefore, the need for the Antidotal and latent states would be put into consideration in this paper giving rise to the Susceptible, Latent, Infected, Recovered and Antidotal (SLIRA) model.

The model proposed in this paper is a modification of the SIRA model of [6] to include computers which are latent

infected by computer viruses in the network. Below is the Piqueira semantic diagram for SIRA computer virus model.

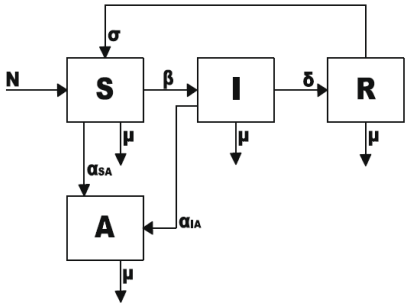


Fig. 1. SIRA model.

Fig.1 Semantic Diagram of SIRA Computer Virus Model

VI. PRELIMINARY ASSUMPTIONS

- The total number of computers on the network (T) can be divided into five classes, namely; Susceptible Computers (S), Latent Infected Computers (L), Infected Computers (I), Recovered Computers (R) and Antidotal Computers (A) with fully functioning computer antivirus software installed on it.

- We assume that the influx rate of computers into the network denoted by *r* and the rate μ of computers leaving the network not due to infection are zero. This implies that $r = \mu = 0$.

- The susceptible computers (*S*) are infected with a rate that is related to the probability of susceptible computers to establish effective communications with infected ones. Therefore, this rate is proportional to the product *SI*, with proportion factor represented by β_1. Or the susceptible computer is converted into latent infected by a proportion do to the product SL and is controlled by α.

- Conversion of susceptible into antidotal is proportional to the product *SA* and is controlled by α_{SA}.

- Latent infected computers may be fixed by means of using antivirus software with the proportion do to the product LA that is controlled by user awareness $u_1(t)$ or it is recovered by the control factor $u_3(t)$. Also, latent infected computers may become infective via the product LI and controlled by the proportion factor β_2.

- Infected computers can be fixed by using anti-virus programs being converted into antidotal ones with a rate proportional to AI, with a proportion factor given by $u_2(t)$ or becomes recovered with a rate controlled by $u_4(t)$.

- Recovered computers can be restored and converted into susceptible with a proportion factor ρ.

From the above preliminary assumptions we move on to generate a semantic diagram of SLIRA model as shown below.

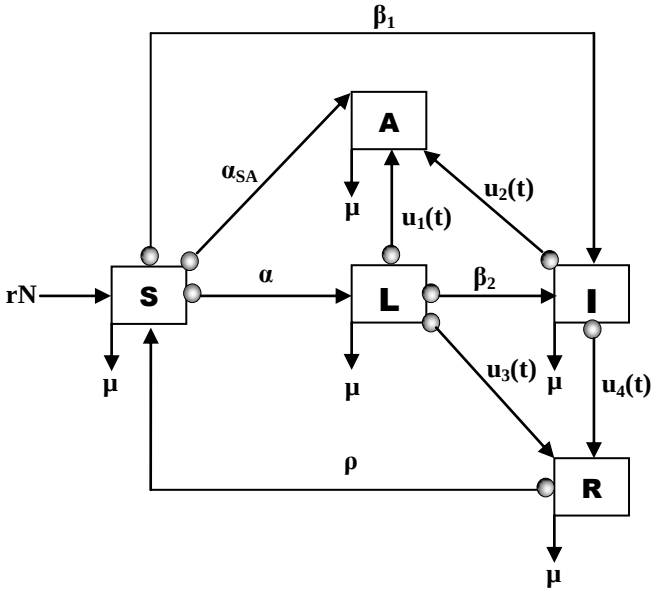


Fig.2 Semantic Diagram of SLIRA Computer Virus Model

The semantic diagram of Fig.2 shows the path and behavior of computer virus and antivirus in a computer network. From the semantic diagram above the following equations can be derived

$$\left.\begin{aligned} S' &= rN - \alpha_{SA}SA - \beta_1 SI - \alpha SL + \rho R - \mu S \\ L' &= \alpha SL - \beta_2 LI - u_1(t)LA - u_3 L - \mu L \\ I' &= \beta_2 LI + \beta_1 SI - u_2(t)IA - u_4(t \\ R' &= u_3(t)L + u_4(t)I - \rho R - \mu R \\ A' &= \alpha_{SA}SA + u_1(t)LA + u_2(t)IA - \mu A \end{aligned}\right\} \quad (1)$$

Under the assumption that then, equations (1) becomes

$$\left.\begin{aligned} S' &= -\alpha_{SA}SA - \beta_1 SI - \alpha SL + \rho R \\ L' &= \alpha SL - \beta_2 LI - u_1(t)LA - u_3 L \\ I' &= \beta_2 LI + \beta_1 SI - u_2(t)I. \\ R' &= u_3(t)L + u_4(t)I - \rho R \\ A' &= \alpha_{SA}SA + u_1(t)LA + u_2(t)IA \end{aligned}\right\} \quad (2)$$

From the provided systems of ODE in equations (2) we have that $\frac{dS}{dt} + \frac{dL}{dt} + \frac{dI}{dt} + \ldots(3)$

On integrating equation (3) it becomes

$S + L + I$ for any instant *t*.

VII. DISEASE-FREE EQUILIBRIUM POINTS

There are disease free equilibrium points, which represent the situations where the infected systems is null (I = 0). These points are given by:

$E_0 = (S = 0, L = 0, I = 0, R = 0, A = T)$ *and*
$E_1 = (S = T, L = 0, I = 0, R = 0, A = 0)$

From the given conditions of E_0 we can generate the Jacobean matrix for the linear system of equation (2). Whereby would be given as:

$$J_{E_0} = \begin{bmatrix} -\alpha_{SA}T & 0 & 0 & \rho & 0 \\ 0 & -(u_1(t)T + u_3(t)) & 0 & 0 & 0 \\ 0 & 0 & -(u_2(t)T + u_4(t)) & 0 & 0 \\ 0 & u_3(t) & u_4(t) & -\rho & 0 \\ \alpha_{SA}T & u_1(t)T & u_2(t)T & 0 & 0 \end{bmatrix}$$

By means of Ruth Hurwitz criterion of Liapunovs stability theory if the corresponding diagonal Eigen values of are it implies that the system of is asymptotically stable. [7]

Since the corresponding eigenvalues of is given as:

$$J_{E_0} = \{\lambda_1 = -\alpha_{SA}T, \quad \lambda_2 = -(u_1(t)T + u_3(t)), \quad \lambda_3 = -(u_2(t)T + u_4(t)), \quad \lambda_4 = -\rho, \lambda_5 = 0\}$$

Respectively.
Thus, the system E_0 is asymptotically stable.

On observing the A-axis we see that it remains constant for any given initial condition. Also, we notice that λ_1 and λ_4 are real and negative which are conditions necessary for attaining asymptotic stability hence we are left to analyze λ_2 and λ_3.

But

$$\lambda_2 = -(u_1(t)T + u_3(t)) \Rightarrow u_1(t)T > -u_3(t) \Rightarrow T > \frac{-u_3(t)}{u_1(t)} \Rightarrow$$

…(4)

$$\lambda_3 = -(u_2(t)T + u_4(t)) \Rightarrow u_2(t)T > -u_4(t) = T > -\frac{u_4(t)}{u_2(t)} \Rightarrow T < \frac{u_4(t)}{u_2(t)}$$

… (5)

and of equations (4) and (5) are the conditions necessary for to be asymptotically stable respectively.

The Jacobean matrix of is given as:

$$J_{E_1} = \begin{bmatrix} 0 & -\alpha T & -\beta_1 T & \rho & -\alpha_{SA}T \\ 0 & -(u_3(t) - \alpha T) & 0 & 0 & 0 \\ 0 & 0 & -(u_4(t) - \beta_1 T) & 0 & 0 \\ 0 & u_3(t) & u_4(t) & -\rho & 0 \\ 0 & 0 & 0 & 0 & \alpha_{SA}T \end{bmatrix}$$

Consequently the Eigen values of is given as:

$$J_{E_1} = \{\lambda_1 = 0, \quad \lambda_2 = -(u_3(t) - \alpha T), \quad \lambda_3 = -(u_4(t) - \beta_1 T), \quad \lambda_4 = -\rho, \lambda_5 = \alpha_{SA}T\}$$

Since not all the Eigen values are real and negative it implies that the system of is not stable by means of Ruth Hurwitz criterion.

VIII. BASAL REPRODUCTION RATE (R_0)

This is a bifurcation parameter meaning that, if $R_0 > 1$, all disease free equilibrium points are unstable and the epidemic process persists. If $R_0 < 1$, there is asymptotically stable disease free equilibrium point; thus, the disease can vanish. In this model, the basal reproduction rate can be determined by analyzing the stability of .

From equations (4) and (5) we obtain our Basal Reproduction Rates (R_{01} and R_{02}) as:

… (6)

… (7)

Consequently if are < 1, the virus propagation within the network is avoided. Then the limiting infection rates of antidotal computers are given by:

… (8)

… (9)

V. ENDEMIC EQUILIBRIUM POINT

Endemic equilibrium point is characterized by the existence of infected systems in the network. This is to say that $I \neq 0$.

$$E_2 = \left\{ S = \frac{u_4(t)}{\beta_1}, L = 0, \; I = \frac{T - \frac{u_4(t)}{\beta_1}}{1 + \frac{u_4(t)}{\rho}}, \quad R = \frac{T - \frac{u_4(t)}{\beta_1}}{1 + \frac{\rho}{u_4(t)}}, \quad A = 0 \right\}$$

The Jacobean matrix of is given as:

$$J_{E_2} = \begin{bmatrix} -\beta_1\left[\frac{T - \frac{u_4(t)}{\beta_1}}{1 + \frac{u_4(t)}{\rho}}\right] & -\frac{u_4(t)}{\beta_1} & u_4(t) & -\alpha_{SA}\frac{u_4(t)}{\beta_1} \\ 0 & \frac{u_4(t)}{\beta_1} - \beta_1\left[\frac{T - \frac{u_4(t)}{\beta_1}}{1 + \frac{u_4(t)}{\rho}}\right] - u_3(t)) & 0 & -u_3(t)\left[\frac{T - \frac{u_4(t)}{\beta_1}}{1 + \frac{u_4(t)}{\rho}}\right] \\ \beta_1\left[\frac{T - \frac{u_4(t)}{\beta_1}}{1 + \frac{u_4(t)}{\rho}}\right] & \beta_1\left[\frac{T - \frac{u_4(t)}{\beta_1}}{1 + \frac{u_4(t)}{\rho}}\right] & -\rho & \alpha_{SA}\frac{u_4(t)}{\beta_1} + u_3(t)\left[\frac{T - \frac{u_4(t)}{\beta_1}}{1 + \frac{u_4(t)}{\rho}}\right] \\ 0 & u_3(t) & 0 & 0 \end{bmatrix}$$

Using MATLAB 7.10.0 (R2010a) to solve for the Characteristic polynomial of E_2 we have the given result below:
NB: $\lambda = \beta_1 * I$

The characteristic polynomial $P(\lambda)$ of the Jacobean matrix is given as:

$$P(\lambda) = \lambda^5 + [U_3(t) + \rho - \alpha*S - \alpha_{sa}*S + \beta_1*I + \beta_2*I - I*U_2(t)]*\lambda^4 + [\rho*(U_3(t) - \alpha*S + \beta_1*I + \beta_2*I) - (\alpha_{sa}*S + I*U_2(t))*(U_3(t) + \rho - \alpha*S + \beta_1*I + \beta_2*I) + \beta_1*I*U_4(t) + \beta_1*I*(U_3(t) - \alpha*S + \beta_2*I)]*\lambda^3$$

+ [ρ*(β1*I*U4(t) + β1*I*(U3(t) - α*S + β2*I)) - (αsa*S + I*U2(t))*(ρ*(U3(t) - α*S + β1*I + β2*I) + β1*I*U4(t) + β1*I*(U3(t) - α*S + β2*I)) - β1^2*I^2*U4(t) - β1*I*U4(t)* ρ+ β1*I*U4(t)*(U3(t) - α*S + β1*I + β2*I)]* λ^2
+ [β1^2*I^2*U4(t)* ρ- ρ*(β1^2*I^2*U4(t) - β1*I*U4(t)*(U3(t) - α*S + β1*I + β2*I)) - (αsa*S + I*U2(t))*(ρ*(β1*I*U4(t) + β1*I*(U3(t) - α*S + β2*I)) - β1^2*I^2*U4(t) - β1*I*U4(t)* ρ+ β1*I*U4(t)*(U3(t) - α*S + β1*I + β2*I)) - β1*I*U4(t)* ρ*(U3(t) - α*S + β1*I + β2*I)]* λ
+ (αsa*S + I*U2(t))*[ρ*(β1^2*I^2*U4(t) - β1*I*U4(t)*(U3(t) - α*S + β1*I + β2*I)] – [β1^2*I^2*U4(t)*ρ+ β1*I*U4(t)* ρ*(U3(t) - α*S + β1*I + β2*I)]

Again with the aid of MATLAB 7.10.0 (R2010a) the Characteristic Roots of the Characteristic Polynomial is given as:

$\lambda_1 = 0$

$\lambda_2 = \alpha_{sa}*S + U_2(t)*I$

$\lambda_3 = - \rho/2 - (\beta_1^2*I^2 - 2* \beta_1*I* \rho - 4*U_4(t)* \beta_1*I + \rho^2)^{(1/2)/2} - (\beta_1*I)/2$

$\lambda_4 = (\beta_1^2*I^2 - 2* \beta_1*I* \rho - 4*U_4(t)* \beta_1*I + \rho^2)^{(1/2)/2} - \rho/2 - (\beta_1*I)/2$

$\lambda_5 = \alpha*S - U_3(t) - \beta_2*I$

IX. GRAPHICAL SOLUTIONS

In this section, the derived equations of equation (2) is simulated by using MATLAB with ODE45 scheme. The values of parameters are as presented below.

Parameter	α_{SA}	α	β_1	β_2	ρ	τ
Value	0.2	0.2	0.6	0.3	0.1	5

Now we consider the initial population containing susceptible nodes $S(0) = 6$, Latently infected nodes $L(0) = 1$, infected nodes $I(0) = 2$, recovered nodes $R(0) = 1$ and antidotal nodes $A(0) = 1$ for numerical simulation

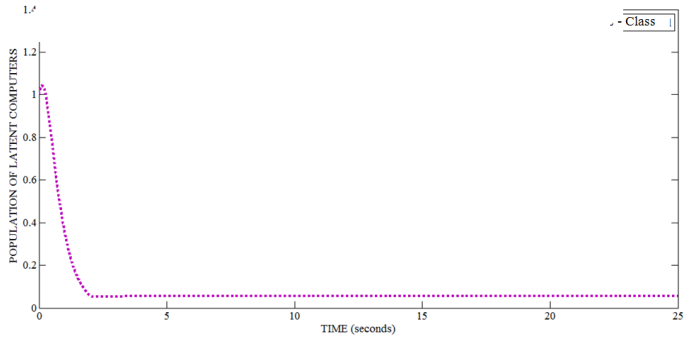


Fig.3 The plot shows the population of latent infected computer against time

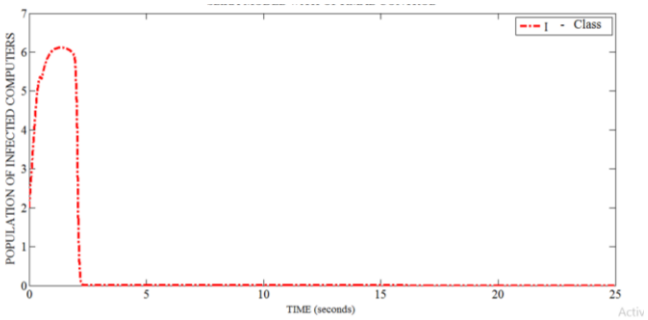


Fig.4 The plot shows the population of infected computer against time

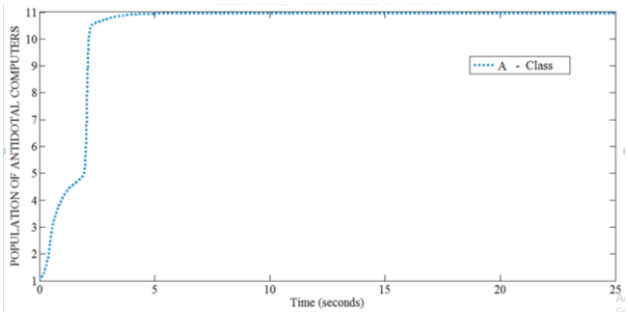


Fig.5 The plot shows the population of antidotal computer against time

VII. DISCUSSION

Considering that P(λ) is of the form:

$$P(\lambda) = a_0\lambda^5 + a_1\lambda^4 + a_2\lambda^3 + a_3\lambda^2 + a_4\lambda + a_5$$

On close examination of P( ) we have that;

- $\lambda_1 = 0$ is a root of P(λ) which implies that 0 is a root of P(λ)
- and 0 is a root of P(λ) whenever $\beta_1 = \rho = u_3(t) = 0$ by observing λ_2, λ_3, λ_4 and λ_5
- Hence, the system of E_2 would be asymptotically stable whenever $\beta_1 = \rho = U_3(t) = 0$
- By means of Ruth Hurwitz criterion, the system of E_2 is unstable as long as $\beta_1 = \rho = U_3(t) > 1$

Also, the graphical solution of fig.3 and fig.5 shows that the virus infection has not been completely eradicated from the network as the graphical simulation of the latent infected computers show.

VIII. CONCLUSION

From the graphical simulation we observed that ten (10) computers of the eleven (11) computers on the network where completely antidotal with functioning limiting infection parameters . while the 11th computer is antidotal but not completely as a result of latent infection on that computer system which explains why is asymptotically stable. This implies that the aim of every organization which has a computer network should be to have every computer fully protected by means of making them antidotal and as well as having computer system administrators who has vast

knowledge of computer virus behavior as this would go a long way in aiding the detection of latent viruses on the network computers as well as stopping them before they become infective. It is also observed in system E_1 where the total population of computers on the network is susceptible that the Jacobean matrix was unstable. This is more or less an indication of the fact that susceptible systems on the network are prone to viral infection since there is no available parameter in the susceptible class of machines for checkmating computer viral infections. Also, the Jacobean system of clearly indicates that as long as the birth rate ( ) of the virus is zero or less the effect of the threshold on the network is negative there by leading to an asymptotically stable system. Hence to eradicate the infection from an infected network the aim should be to drive the birth rate of the virus to negative. More importantly to note is the fact that the latent infection state cannot be neglected in any mathematical model of computer virus spread since as shown in our graphical simulation above, the latent computer on the network never attains zero (0) which implies that if the necessary conditions are in place there could be another round of virus spread on the network.

However, it should be noted that driving the birth rate of computer viruses to negative is an effective means of deterring the spread of the computer virus on the network systems. This does not go to imply the non-existence of computer virus on the network system; a consequence of the fact that even the most effective Antivirus Software can not accurately detect the presence of new and modified computer viruses, a fundamental flaw related to the fact that most antivirus software's detect computer viruses based on the computer virus definitions that they have within their viral scan mechanism.

REFERENCES

[1] Frederick B. Cohen (1990); A short Course on Computer Viruses, ASP Press, Pittsburgh, USA.

[2] Denning P.J., (1990); 'Computer Under Attack', Addison-Wesley, Reading, Mass.

[3] Kephart J.O., White, and S.R. (1991): Directed Graph Epidemiological models of computer viruses. In: IEEE Symposium on security and privacy. pg343-361.

[4] Giuseppe Serazzi and Stefano Zanero (2004); 'Computer Virus Propagation Models', Dipartmento di Elettronica e Informazione, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy

[5] Stanford S., V Paxson, V. Weaver, N. (2002); "How to Own the Internet in Your Spear Time", in Proceedings of the 11th USENIX Security Symposium (Security' 02).

[6] Piqueira J.R.C., Navarro B.F. (2005); "Epidemiological Models of computer viruses", IEEE Symposium on Security and Privacy.

[7] K. Ogata (1997). "Modern Control Engineering", Prentice Hall, New Jersey

An Hybrid K-Means and Discrete Wavelet Algorithm For Solving Identity Fraud in Automated Teller Machine (ATM)

Simbiat O. Salawu
Computer Science Department,
University of Ilorin
Ilorin, Nigeria
simbiatoladoja@rocketmail.com

B. A. Oluwade
Computer Science Department,
University of Ilorin
Ilorin, Nigeria
info@dewadeconsult.com

Moshood A. Hambali *
Computer Science Department
Federal University Wukari
Wukari, Nigeria
hambali@fuwukari.edu.ng

Morufat D. Gbolagade
Physical Science Dept.
(Computer Science program)
Al-Hikmah University
Ilorin, Nigeria
dammyconsult@gmail.com

*Corresponding Author

Abstract — One of the security issues in Nigeria is the proliferation of identity theft and fraud relating to the uses of automated teller machine (ATM) from banks. Research shows that perpetrators of this act are hackers that make use of the last four digits of the sixteen digits printed at the front of the ATM card (ATM card number), the expiry date and card verification value (CVV) number printed at the back of the ATM card. Fingerprint contains unique features for individuals which can serve as a backbone security for everyone. There are several fingerprint point matching, but the researchers deal mainly on minutiae fingerprint point matching and how problems associated with it can be overcome. However, many researchers have worked on minutiae but the problem posed by minutiae is that it causes noise as a result of dirt, injured finger and many other obstacles thereby bringing about low quality images as the result. This paper therefore proposes a framework that helps solve identity fraud with ATMs using hybridized k-means and discrete wavelet algorithms for segmentation, feature extraction and matching. Front-end of the application was implemented using Java programming language and MySQL database (wampserver) was also used for back-end.

Keywords— K-means algorithm, Discrete wavelet algorithm, Hackers, Automated Teller Machine (ATM), Fingerprint.

X. INTRODUCTION

A prominent security issue in Nigeria is the increase in identity theft and fraud in relating to the usage of automated teller machine (ATM) from banks [1]. Research shows that perpetrators of this act are hackers that make use of the last four digits of the sixteen digits printed on the ATM card, this is tagged as the ATM card number, the expiry date of the ATM card is one of the information they work on, with the card verification value (CVV) number written at the back of the ATMs. The hackers make use of software to enhance their work, they break into bank accounts and transfer fund from one account to another, they sometimes steal peoples ATM card and tries to break the security by generating a code different from the owner's personal identification number (PIN code) thereby making withdrawals from the person's account without the person's consent.

Automatic validating someone's identity using his face, iris or fingerprint is no more science fiction, but rather it has become a normal daily routine of authenticating procedure in many places. Biometric recognition have emerged as most promising option for securing valuable information based on individual's physical appearance and/or behavioral features [2]. For the fact that individual's natural traits/features cannot be forgotten, forged, misplaced or stolen. Biometric-based technologies can be group into two: identification of physiological features which includes face, fingerprint, finger geometry, hand geometry, hand veins, ear, palm, iris, retina, voice and so on. Second one is behavioral traits which includes gait, keystroke dynamics, DNA, signature and so on. [3], [4], [2]. For numerous reasons, the fingerprint is generally acceptable to be the most practical features. Fingerprints recognition are easily accessible, requires least effort from the user, capture only information required for the recognition process and provides relatively good performance. Another motivation for its acceptance is the relatively cheap price of fingerprint sensors [5], [6]. There are five categories of fingerprint based on the grouping of their pattern types, they are: arches, tented arches, left loops, right loops and whorls. Most fingerprint verification system make uses of minutiae matching point, minutiae points are the regions in a fingerprint image where the fingerprints splits into two new ridges [7]. Minutiae is believed to be more effective and ascertained method in terms of cost, time and man power [8], but the major problem posed by minutiae is that, it is prone to noise as a result of dirt or any other obstacles.

K-means algorithm is used mainly for clustering [9], in the proposed system k-means algorithm is used for

segmentation, and the purpose of segmentation is to eliminate the background region that could cause noise as a result of dirt, and to obtain an obstacle free foreground region as the region of interest (ROI). Therefore, in order not to crop out false features and to minimize processing time prior to segmentation, this work used Improved Segmentation method based on K-means for sensor Interoperability (ISKI) method [10]. Wavelet on the other hand is used to extract features from the segmented ROI and then stored into database for reference, comparison and matching purpose for a genuine authentication process. Java programming language is used for front-end because of its platform independence, scalability, easy integration, high level of security, dynamism, easy implementation and upgrade. It follows that if this system is properly implemented on ATMs, the level of insecurity will be reduced to the barest minimum if not eradicated totally.

II. STATEMENT OF PROBLEM

The technology advancement in information technology has made transaction possible anywhere, and any point in time. Online banking has made banking activities to be fraud prone, and thereby giving room for fraudulent acts by hackers and internet thieves to gain access to others account information. It follows that this terrible acts cause threats to life and properties of the account holders.

It is a known fact that hackers' activities have negative impact on the nation's economic development and infrastructure [11]. Minutiae points are the regions in a fingerprint image where the fingerprint splits into two new ridges [7]. Minutiae is also believed to be more effective and ascertained method in terms of cost, time and man power [8] but the major problem posed by minutiae is that it is prone to noise as a result of dirt or any other obstacle thereby bringing about low quality images as a result.

This research therefore seeks a more secure way for transaction processing using automated teller machine (ATM); this is achieved by proposing a system that implements K-means and discrete wavelet algorithm for the process of segmentation, feature extraction and matching; so as to obtain a more secure way of authentication before any transaction can be made from the ATM.

III. LITERATURE REVIEW

Biometric system has become most acceptable mean of recognition in the recent times, the process compare the incoming information with the one in the database and confirming if the information tallies i.e. authentication process (verification and identification process) [12].

Sophisticated processing in biometric system can be taught of, as individual identify a specific person, and when someone presented to the system, it can decides whether the person is someone familiar or not [12]. The algorithm used in biometric also permit the matching (compare) of an enrolled (existing) template with a new template just created for the validation of an identity, termed a live template [13]. When an existing template and a live template are compare, the system estimates the closeness of the two templates. If the match is close enough, a person will be confirmed and authenticated. If not, he will not be verified and deny access to the system.

IV. REVIEW OF EXISTING RELATED RESEARCH

The use of ATM which has been embraced by most individuals still leaves them with fear of security which needs to be tackled with all seriousness it deserves, before customers can be comfortable with the use of ATM for all their transactions.

ATM cards must be very secure to the extent that, even if the owner misplace or lost the card, and it will still have confidence that the hacker or attacker will be unable to use the card. Since security measures at ATM points play a momentous role in averting attacker or hackers from making illegal withdrawals or transactions on customer's money, several researches have proposed the used of fingerprint biometrics on ATMs before withdrawal or any form of transaction can be done. Jeroen, Ileana, Koen and Emile [14] provided adequate information about the benefits and limitation of incorporating biometrics in a PIN-base payment authentication system. Based on their review they proposed a biometric method that can be incorporated into a PIN-based verification infrastructure by binding a fixed token, renewable number to a noisy biometric sample. The South African Social Security Agency (SASSA) has introduced a new SASSA Payment Card that has a fingerprint authenticated features. The card is a MasterCard SASSA-branded smart payment, which has an embedded chip containing personal information, fingerprint and secret PIN. With the card, card holders can easily withdraw and make payment at point-of-sale (POS) center, purchase airtime, pay water and electricity bills, or open accounts [15]. Wang et al. [16] proposed a fingerprint orientation model based on 2D Fourier expansions (FOMFE) in the phase plane. However, their proposed FOMFE does not necessitate prior information of singular points, it is able to define the overall ridge topology without flaw. Fengling et al. [17] proposed an encryption/authentication smartcard based scheme for ATM banking system. Their proposed smartcard has two stages of security check; the first layer performs authentication based on information embedded on the smartcard. At the second layer, fingerprint authentication is done via feature and minutiae matching. Das and Jhunu [18] discussed on susceptibilities and the increasing in criminal activities occurring at ATMs and offered a fingerprint authentication prototype for enhancing security of ATM card. The systems assumed the same measure as the current work by framing modules for enrolment of fingerprint, fingerprint images enhancement, feature extraction, database and matching. Santhi and Kumar [19] provide a review detail on various existing biometric systems and also identified the strengths and limitations of each of the system presented. Hence, they proposed Personal Identification Image (PII) security enhancement method to secure ATM. Bhosale and

Sawant [20] and Ibiyemi et al. [21] presented a new hybridize feature-based biometrics models which replaced card system with biometric technology. Their model used fingerprint, iris and PIN to provide reliable and full-proof ATM authentication.

Mali et al. [22] presented a framework for real time secured ATM application using a combination of thumb print, face recognition and PIN. Their proposed framework is expected to register thumb print and face features and store at a server in encrypted format. Validation is done by decrypting the stored patterns in database, and compare with input pattern before access privilege is granted for ATM operations. Their system employs Principal Component Analysis (PCA) and Eigen algorithm for the face recognition, LSB algorithm for stenography and AES algorithm for cryptography. However, the framework seems promising, but detailed implementation and evaluation is not presented in their work. Abayomi et al. [23] proposed an enhanced e-banking system where customer can access multiple accounts over different banks institutions with a single ATM card as fingerprint authentication. A match-on-card technique was used, which relies on a one-to-one matching where the data from the ATM fingerprint sensor is compared only to the template stored on the user's ATM card. This will help in privacy concern of users; the system will also help the users to have access to multiple accounts with a single ATM card. It is secured and help in reducing ATM fraud. The paper employed the unique features of fingerprint to overcome the flaw of the PIN based ATM authentication. Selina and Jane [24], proposed a method of existing security of the ATM system with integration of fingerprint of the user into the bank's database, so as to serve as additional means of authenticating the user. This was accomplished by modelling and constructing an ATM simulator that imitate a typical ATM system. Krishnamurthy and Redddy [25], are of the opinion that the working of ATM machine can base on when customer place finger on the fingerprint module of ATM machine; ATM will automatically generates 4-digit code message and send it to the mobile phone of the authorized customer via GSM modem connected to the microcontroller. The received code should be input into the system through the touch screen keys of ATM. After inputting, it verifies whether it is a valid one or not and grants the customer access to perform whatever transaction it wants. The problem with this approach is that, if network of mobile network provider failed, the user may not be able to receive code message send to his/her mobile phone. Jimoh and Babatunde [26], proposed a Short Message Service (SMS) Verification enhancement for the ATM Authentication. The usability of the developed prototype was verified using heuristic evaluation method with the help of a questionnaire. The report from ten (10) respondents who are ATM users in the country and the data obtained were analyzed using Statistical Package for Social Science (SPSS). Charles and Wilfred [27], presented a unified (single) smart card-based ATM card with biometric enhanced cash dispenser for all banking transactions. Their proposed work is expect to reduce the number of ATM cards carried by an individual and the biometric integrated on ATM is to increase the level of security compare to the PIN based which are currently being used.

However, with the advent of the proposed cloud computing the level of insecurity and ATM crime rate will drastically increase as fraudsters and hackers will have more access to banking systems, information and tools, the researcher therefore, proposes a more secure way of dealing with ATM fraud by providing a novel way of biometric authentication by hybridizing two popular algorithms for the proposed system so as to prevent hackers and ATM fraudsters from withdrawing money illegally without the account owner's consent. It follows that if the proposed system is properly implemented the level of insecurity in ATM dealings will drastically reduce if not totally eradicated.

V. RESEARCH METHODOLOGY

The proposed framework for the ATM Fingerprint Recognition process is presented (Fig. 1), starting with feature extraction and matching. The framework is adapted from the conventional biometric recognition process. In this system, for each training fingerprint image, k-means algorithm is used for the pre-processing step to perform fingerprint segmentation; a center region (core point) area of the fingerprint is detected using discrete wavelet transform (DWT) to crop the region of interest (ROI). The fingerprint features are extracted using DWT and stored into the database. During the testing phase, features will be extracted from the scanned fingerprint image from the scanner and matched with the corresponding features stored earlier. Distance based matching is used for similarity measure.

VI. FINGERPRINT PRE-PROCESSING

All biometric recognition systems need to filter the live sample in some way in order to find the relevant parts. That is, by removing the background from a captured image or by transforming a fingerprint scan from gray scale to black and white; or transform the lines to a width of just one pixel, all this are done at the pre-processing stage of the recognition system. Fingerprint pre-processing consists of four major steps which includes; image alignment, segmentation, enhancement, binarization and thining. All these process helps to improve the feature extraction and matching of fingerprint templates [28].

Fingerprint segmentation is an important section in automatic fingerprint recognition system. Effective segmentation increase the detection accuracy of minutiae and subsequent reduce the processing time, thus it augment the performance of the entire system. In this research work, K-means algorithm is used for fingerprint segmentation at the pre-processing stage.

Clustering is an unsupervised machine learning process which classify set of physical or abstract objects into similar objects or groups [29]. Fingerprint segmentation can be

viewed as a Bi-class clustering task, whose aim at differentiate the foreground cluster (that is, fingerprint) from the background one. Consequently, segmentation are performed on individual fingerprint image by using clustering algorithm without pre-label information on the image. The researchers' approach employs the Improved Segmentation method based on K-means for sensor Interoperability (ISKI) for fingerprint segmentation algorithm [30].

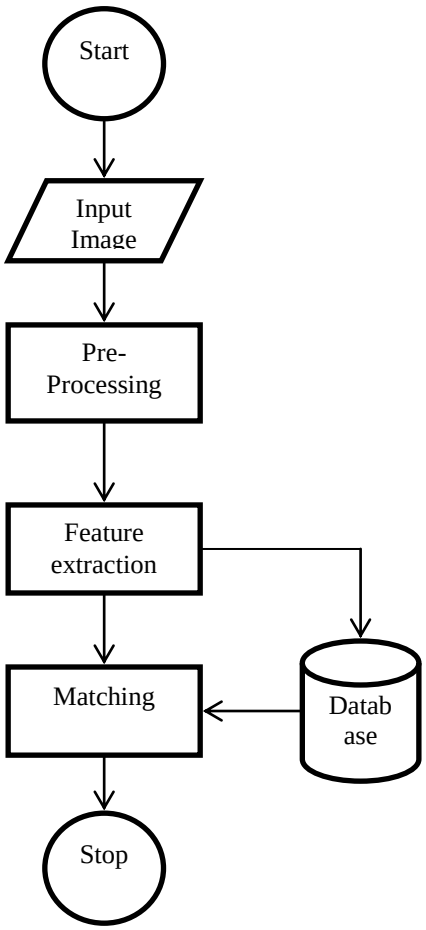


Fig. 1: Proposed framework for Fingerprint Recognition

K-means algorithm was proposed by [9] is a famous employed clustering method. In [10] a k-means based segmentation method named SKI (Segmentation based on K-means for sensor Interoperability) is proposed to solve the problem of sensor interoperability. In SKI, each fingerprint image is grouped into non-overlapping chunks (blocks) with the same size of w×w pixels, and three block-wise features: coherence, mean and variance (CMV) are extracted as the feature vectors. For individual fingerprint image, k-means algorithm is employed to cluster the blocks into two clusters. A classification pre-processing is done to decide which cluster designates for the foreground blocks since the outputs of k-means are of two clusters without indicating foreground or background. Thereafter, SKI applies morphological post-processing to remove the noisy blocks.

With the aim of effectively differentiate the blocks around the border as foreground or background, by following the earlier work [10], the researchers proposed an improved k-means based sensor interoperable segmentation (ISKI). For individual fingerprint image, k-means algorithm is firstly employed to cluster the blocks into foreground and background. Then the blocks which have the same distances with the two cluster centers, the process will be repeated again. The fig. 2 depicts the framework of ISKI and the steps involve in ISKI are follow:

Step 1: Each fingerprint is grouped into non-overlapping chunks with the identical size of w×w pixels, and the block-wise CMV features are extracted [10] to represent the feature vectors. The CMV values are normalized using Min-Max normalization into [0, 1].

Step 2: Clustering of the blocks into the foreground and the background cluster is done using K-means algorithm. Firstly, the cluster number is set to be 2. The focal point of the fingerprint image is marked as the cluster center of the foreground cluster, and the edge of the fingerprint image is considered as the cluster center of the background one. Then, each block is allotted to point to the cluster center which has the nearest Euclidean distance to it, and after that, the new cluster centers are recalculated, this procedure is repeated until the value remain constant. When k-means clustering is done, the two cluster centers of each block with distances are documented respectively. An ensign is used to designate if a block is either a foreground block or a background block based on the result of k-means algorithm. It is labeled ensign1.

Step 3: In this step, the blocks with the same distances with two cluster centers are identify and secondary determination is done on these blocks. In each block, the difference value of the two distances from cluster centers is computed. If the value is greater than the threshold Tv (Tv is an empirical parameter), the block belongs to foreground region or background region is determined by the k-means algorithm. Else the block is marked. Secondary determination is conducted on the marked blocks using the following three steps:

Step 3a: Extract a new block-wise feature: called combination of variance and its gradient. This feature is defined to be [31]:

$$VarG = VI(x,y) \times DVI(x,y) \quad \text{Eq.1}$$

$$DVI(x,) = \sqrt{\left(\frac{\partial VI^2}{\partial x}\right) + \left(\frac{\partial VI^2}{\partial y}\right)} \quad \text{Eq.2}$$

$$\frac{\partial VI^2}{\partial x}(x,y) \approx \frac{VI(x+1,y) - VI(x-1,y)}{2} \quad \text{Eq.3}$$

$$\frac{\partial VI^2}{\partial y}(x,y) \approx \frac{VI(x,y+1)-VI(x,y-1)}{2} \quad \text{Eq.4}$$

where VI is the variance of the block,
DVI gradient of the block,
x, y is the coordinate distance from the block center,
and VarG is the combination of variance and its gradient.

For every marked block, the feature is extracted and normalized into [0, 1] using Min-Max normalization. If this feature value is more than the threshold Tv, that block is considered as foreground block, else it is considered as background block. The ensign that specifies a marked block as a foreground block or a background block based on the decision of Step 3a is called as ensign2. Tv is an adaptive threshold which is equivalent to the median value of the combination features.

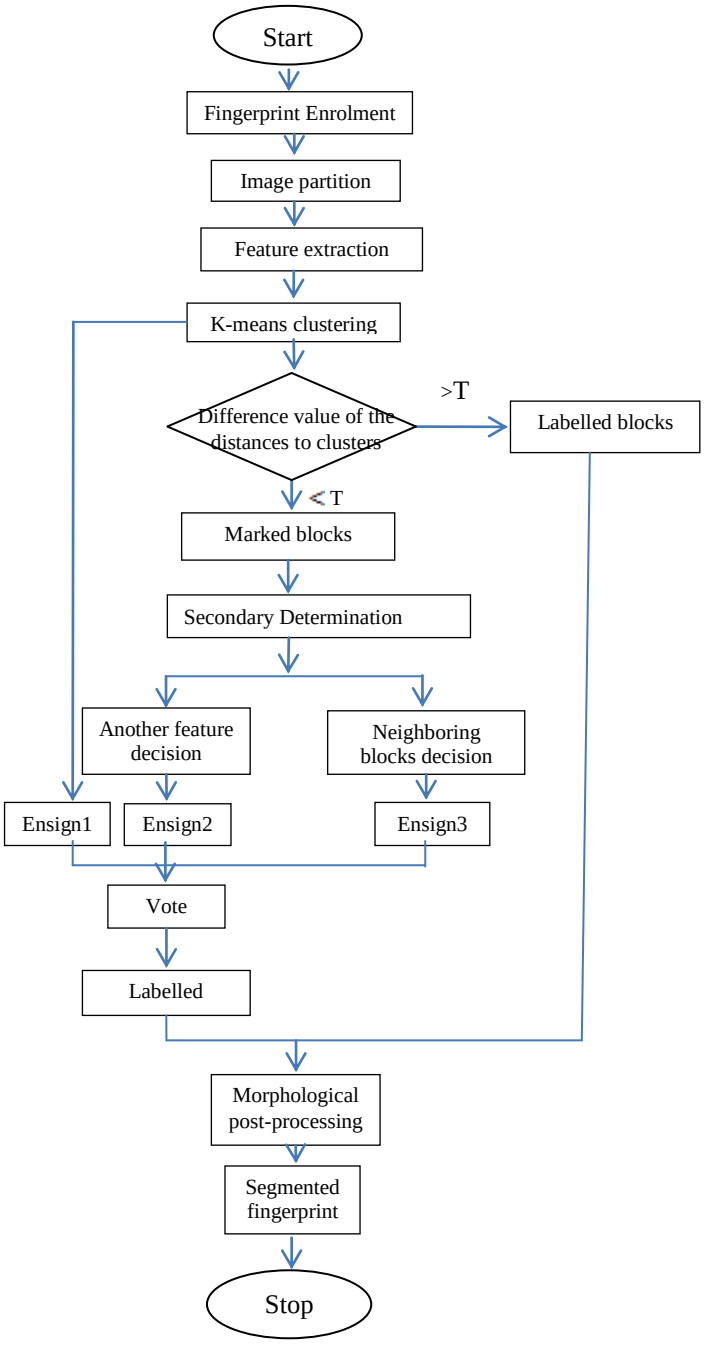


Fig. 2: Framework of ISKI

Step 3b: Let consider the results of k-means of 8 neighborhoods marked blocks. If more than 4 neighborhoods are considered to be foreground blocks by k-means clustering, the block is designated as foreground block, else the block is designated as background block. The ensign that specifies a marked block as a foreground block or a background block based on the decision of Step 3b is called as ensign3.

Step 3c: Vote to decide the label of the block. For every marked block, three steps have been used to decide if it is a

foreground or background block. The results of the three steps are recorded by ensign1, ensign2 and ensign3. The recorded data in the ensigns will be used to decide the block that belongs to the foreground area or background area by voting. If marked block has vote that greater than two votes as foreground (or background) block, the marked block is labelled as foreground (or background) block.

Step 4: Noisy blocks are eliminated by Morphological post-processing.

A. Discrete Wavelet Transform (DWT)

Wavelet transform (WT) denotes image as a summation of wavelets on various resolution levels. The strong point of the WT is that it provides high temporal localization for high frequencies while offers better frequency resolution for low frequencies. Therefore, WT is one of good algorithms to extract local features of the image. The hierarchical wavelet transform employs predecessor wavelet functions and its associated scaling functions to breakdown the original signal/image into various sub bands. The decomposition process is repeatedly applied on successor bands (sub bands) to produce the subsequent level of the hierarchy.

This illustrations one level DWT. An image f(x, y) whose forward discrete transform can be defined in the form of following general relation:

where, i = {H, V, D} and j_0 is an arbitrary initial scale.

$W\varphi(j_0, m, n)$ coefficients express the approximation details of image f(x, y) at scale j_0.

$W_i\varphi(j_0, m, n)$ coefficients add horizontal, vertical and diagonal details of image f(x, y) at scale j_0. In this research work, we applied wavelets on the segmented image which is the output of our pre-processing stage and the features was extracted from the cropped image around the Core Point. In this work, we used level 2 daubechies transform. Daubechies is employed to solve the problems associated with JPEG compression and random additive noise.

B. Feature Extraction

Feature extraction is dealt with the quantification of texture characteristics in terms of a collection of descriptors or quantitative feature measurements, usually called feature vector. DWT Feature extraction commences extraction by transform the input image into a 2-level discrete wavelet transform decomposition. At each level, the WT breakdown the given image into three directional paths, that is, horizontal, diagonal and vertical detail sub bands in the direction of 0, 45 and 135 respectively apart from the approximation (or) smooth sub band. Standard deviation and the energy based techniques are applied independently on individual sub-band information. The standard deviation is then estimated from the horizontal, vertical and diagonal details referred to as {dj_1 dj_2 dj_3}.the equation 5 is used for estimating standard deviation:

$$\sigma_k = \frac{1}{M \times N} \sum_{I=1}^{N} \sum_{j=1}^{N} E[W_k(i,j) - \mu_K] \qquad \text{Eq.5}$$

where, $W_k(i, j)$ is the k^{th} wavelet decomposed sub-band. M x N is the wavelet decomposed sub-band. μ_k is the mean value of k^{th} decomposed sub-band. Similarly Eq.6 used to estimate energy function.

$$Energy = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |X(m,n)| \qquad \text{Eq.6}$$

where X (m, n) is a discrete function whose energy is to be calculated.

C. Matching

Distance based classifier is used in the proposed system for fingerprint recognition. Relative distances of individual Feature Vector are compared with the already stored Feature Vectors. Euclidean distance metric given in Eq.7 is used to calculate the similarity or match value for given pair of features. Zero distance indicates a perfect match, and feature tends towards mismatch as the distance increases.

$$D_{(x,y)}^{Eucli} = \sqrt{\sum_{i=0}^{N} (x_i - y_i)^2} \qquad \text{Eq.7}$$

VII. RESULTS AND DISCUSSION

A. The Admin Login

Admin login (fig. 3) allows the admin to login and prevents all unauthorized users from logging in or access the database so as to have a secured environment and smooth operation for the process of extraction through the normalization of image using Min-Max normalization and matching using the designed threshold (Tv) by the researcher or retrieval of data as the case may be.

Fig. 3: Admin login page

B. The Enrolment Phase

The enrolment phase (fig. 4) involves capturing of the account holder's bio-data and all necessary information, assigning account number to each enrollee and taking the biometric details also, the next of kin's information is contained in the data captured, this is then stored in the database for further reuse. The biometric fingerprint has four features; the first one is the biometric image, binarized image, thinned image and the minutiae image. Once all the necessary information need has being captured and the passport of the

account holder has being loaded then the system displays enrolment **SUCCESS** otherwise **ENROLMENT NOT SUCCESS.** The captured information is subject to editing as the information can be edited from time to time if need arises. It must be noted that the editing can only be done by an authorized person.

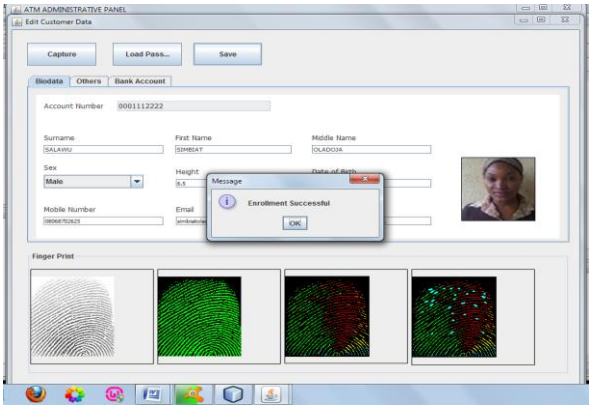


Fig. 4: Enrolment page

C. Database Storage Phase

The database storage houses all information captured from all account holders, can be retrieved when needed and can be edited when the need arises as it is possible for account holders to have a changed information from time to time and the biometrics too changes as the individual grows so the biometric details too should be updated from time to time. The account number, passport and biometrics stored in the database helps verification process and make it easy.

D. System Verification Phase

In the verification phase an account number, and biometric fingerprint has to input into system, and the system checks if the information stored earlier is matches with the one provided by the customer with the account number and the biometric fingerprint. Then, if it matches, the system displays **VERIFICATION IS SUCCESSFUL** else **NOT SUCCESSFUL** is displayed. The system compare the result gotten from the captured fingerprint with the threshold that has being set by the Admin, if the score gotten is greater than the threshold (Tv) score, the system proceeds to the next phase. If otherwise the system displays **WRONG USER. (**See fig. 5)

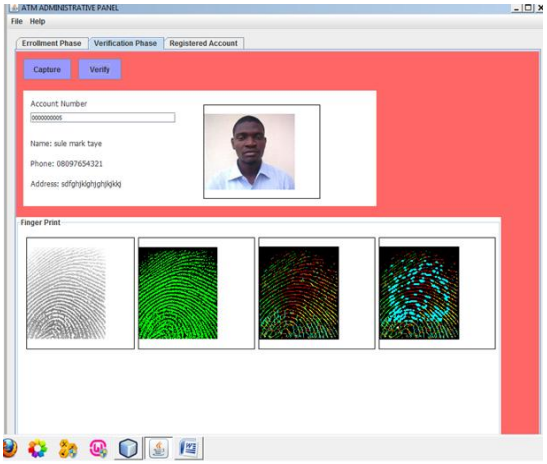
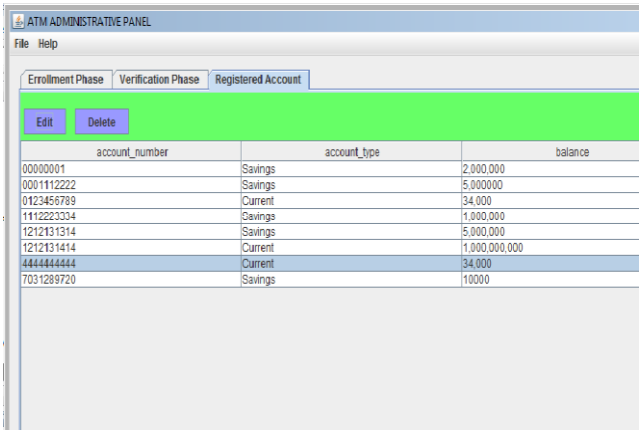


Fig. 5: system verification

E. Registered Account phase

The registered account phase has all the account holders' account numbers details which include information about account balance and account type i.e. saving or current in a tabular form. In this phase, the admin has the privilege to edit the account information and add or updates as the need be. (Fig. 6)



account_number	account_type	balance
00000001	Savings	2,000,000
0001112222	Savings	5,000000
0123456789	Current	34,000
1112223334	Savings	1,000,000
1212131314	Savings	5,000,000
1212131414	Current	1,000,000,000
4444444444	Current	34,000
7031289720	Savings	10000

Fig. 6: Registered Accounts

F. ATM Client phase

The ATM client phase allows users to perform all transactions they want. Once the fingerprint of the user is verified, the user clicks on proceed and the screen displays balance enquiry, withdrawal and others. The account holder can choose the transaction that he/she wants to perform. For instance if the account holder chooses to check **balance** he/she clicks on **Balance Enquiry** then select account type either savings or current and the balance of account is displayed. If he/she chooses to withdraw, he/she will click on withdrawal, the system then prompts the user to choose the account type after which varieties of denominations are displayed (1,000 2,000 5,000 10,000 15,000 20,000 25,000 and others) the user then chooses the amount he/she wishes to withdraw, if the amount is greater than 25,000 then the user selects others and

punch in the amount he/she wishes to withdraw. (Fig. 7 to Fig. 12)

ATM Client page

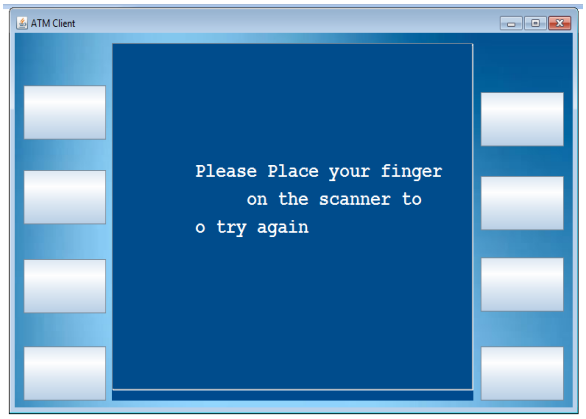


Fig. 7: ATM biometric Fingerprint Authentication page

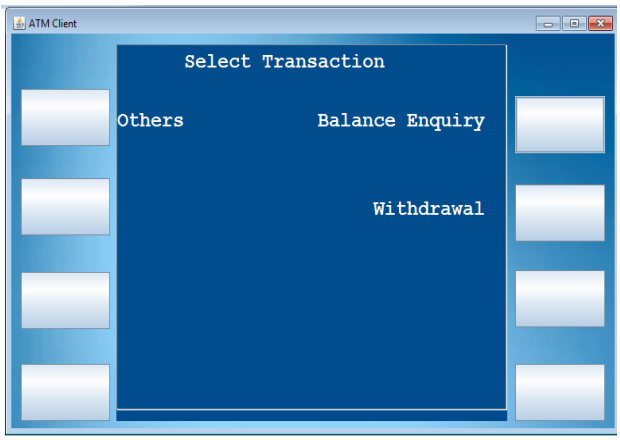


Fig. 8: Transaction Selection Page

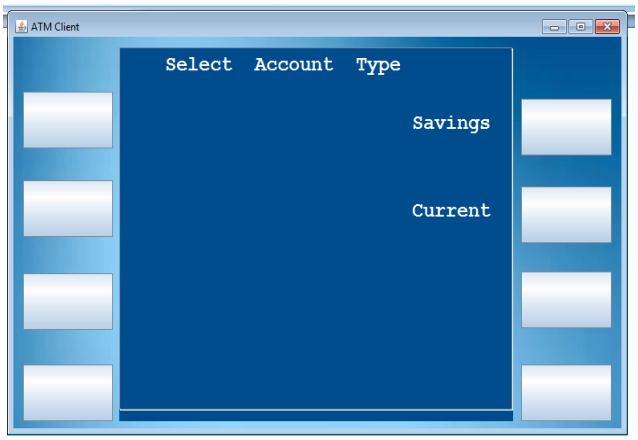


Fig. 9: Account Type Selection Page

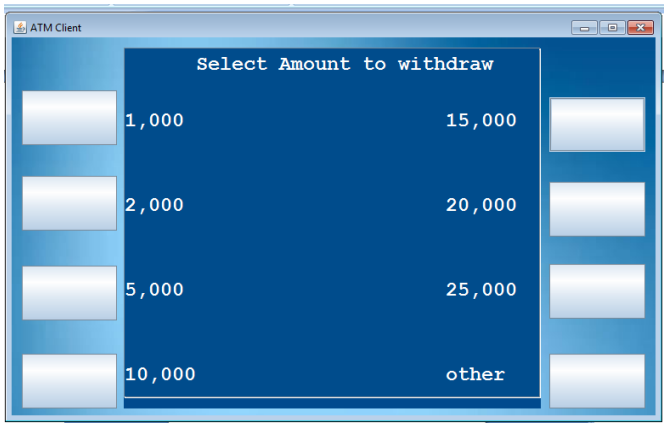


Fig. 10: Withdrawal Amount selection Page

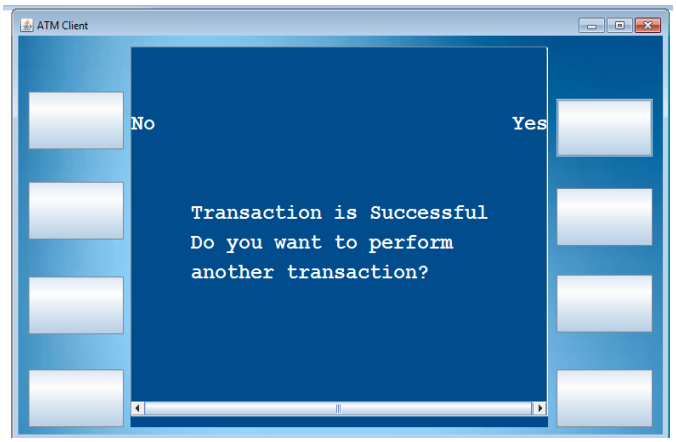


Fig. 11: Transaction Completion Notification Page

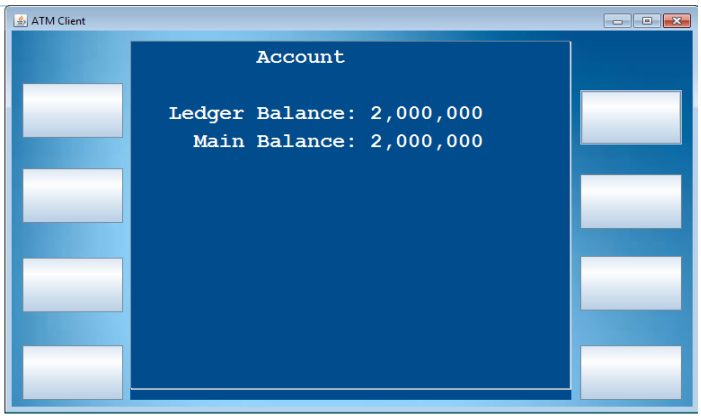


Fig. 12: Account Balance Display Page

VIII. CONCLUSION

This work has gone a long way to check the high-level of ATM fraud in Nigeria as the account holder will have to proof that he/she is actually the person claims to be. They will have to go through some verification processes by supplying their PIN number, some account details and their biometric fingerprints will have to be verified. If it actually tallies with the one in the database before they can initiate any transaction.

With this development, if the system is properly implemented the level of ATM frauds will be reduced drastically in Nigeria if not totally eradicated.

VII. AREAS OF FUTURE WORK

The researchers intend to further the research work on using multimodal biometric i.e., the use of multiple biometric methods such as iris, fingerprint, retina, facial recognition and so on for authentication of account holders before initiating any transaction so as to be sure that they are the true owner of the account they claimed.

REFERENCES

[1] L. A. Patrick, M. O. Ebele and O. Chinedu, "Terrorist Activities and Economic Development in Nigeria: An Early Warning Signal," OIDA International Journal of Sustainable Development, 5(4), pp. 69-78, 2012.

[2] M. A. Hambali and R. G. Jimoh, "Performance Evaluation of Principal Component Analysis and Independent Component Analysis Algorithms for Facial Recognition," Journal of Advances in Scientific Research & Its Application (JASRA), 2, Pg. 47 – 62, 2015.

[3] A. K. Jain, R. Bolle, and S. Pankanti, "Biometrics Personal Identification in Networked Security," A.K. Jain, R. Bolle, and S. Pankanti, Eds.: Kluwer Academic Publishers, 1999.

[4] J. Rabia and A. R. Hamid, "A Survey of Face Recognition Techniques," Journal of International Processing System, 5(2), pp. 41- 68, 2009.

[5] M. B. Asker, "Fingerprint Identification - Feature Extraction, Matching and Database Search," 2002. Retrieved March 2nd, 2013 from http://www.doc.utwente.nl/649431/1/Bazen02_thesis_Fingerprint_Matching_Feature_Extraction,_and_Database_Search.pdf

[6] Y. K. Saheed, J. Aliyu, M. A. Hambali and I. O. Aderibigbe, "A Framework for Examination Clearance Using Biometrics System," Multidisciplinary Innovations & Technology Transfer iSTEAMS (MINTT) Conference - UNIBEN 2015, Book of Proceedings. Prof. S. C. Chiemeke et al. (Eds), 8, 107 – 116, 2015.

[7] M. Eriksson, "Fingerprint Based Identity Verification," Umeå University Department of Computing Science Griaule Biometric, Technical Report, pp. 24, 2001.

[8] R. G. Jimoh, S. O. Abdulsalam, and K. S. Adewole, "Adoption of Fingerprinting as an Automated Access Control Technique in University Hostels," Journal of Systems and Software, 1(4), 2011.

[9] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297, 1967 University of California Press, Ca, USA.

[10] G. P. Yang, G. T. Zhou, Y. L. Yin, and X. K. Yang, "K-Means Based Fingerprint Segmentation with Sensor Interoperability," EURASIP Journal on Advances in Signal Processing, 2010.

[11] HIMSS Privacy & Security Task Force, "Creating a Trusted Environment: Reducing the Threat of Medical Identity Theft," Healthcare Information and Management Systems Society, 2012.

[12] S. Smith, "What is Biometrics Recognition," 2013. Retrieved July 21st, 2013, from http://www.wisegeek.com/what-is-biometrics-recognition.htm.

[13] R. Paul, "Biometrics for Network Security," Prentice Hall Professional, New Jersey, USA, 2004.

[14] B. Jeroen, B. Ileana, D. G. Koen, and K. Emile, "Evaluation of a Template Protection Approach to Integrate Fingerprint Biometrics in a PIN-Based Payment Infrastructure," Netherlands, 2011.

[15] SASSA MasterCard Debit Card Grows Financial inclusion in South Africa, 2013. Retrieved May 23rd, 2015 from: http://newsroom.mastercard.com/press-releases/sassa-mastercard-debit-card-grows-financial-inclusion-in-south-africa/

[16] Y. Wang, J. Hu and D. Phillips, "A Fingerprint Orientation Model Based on 2D Fourier Expansion (FOMFE) and its Application to Singular-Point Detection and Fingerprint Indexing," IEEE Trans Pattern Anal Mach Intell 29: pp. 573-585, 2007.

[17] H. Fengling, H. Jiankun, Y. Xinhuo, F. Yong, and Z. Jie, "A Novel Hybrid Crypto Biometric Authentication Scheme for ATM Based Banking Applications," Lecture Notes Computer Science 3832: pp. 675-681, 2005.

[18] D. Das and S. S. Jhunu, "Designing a Biometric Strategy (Fingerprint) Measure for Enhancing ATM Security in Indian E-Banking System," International Journal of Information and Communication Technology Research pp. 197-203, 2011.

[19] B. Santhi, and R. K. Kumar, Novel Hybrid Technology in ATM Security Using Biometrics," Journal of Theoretical and Applied Information Technology 37, pp. 217-223, 2012.

[20] S. T. Bhosale and B. S. Sawant, "Security in e-Banking via Card-less Biometric ATMS," International Journal of Advanced Technology & Engineering Research 2, pp. 9-12, 2012.

[21] T. S. Ibiyemi, S. E. Obaje, and J. Badejo, "Development of Iris and Fingerprint Biometric Authenticated Smart ATM Device & Card," 24th National Conference of the Nigeria Computer Society (NCS), 2012.

[22] P. Mali, S. Salunke, R. Mane, and P. Khatavkar, "Multilevel ATM Security Based on Two Factor Biometrics," IJERT 1, pp. 8, 2012.

[23] A. Abayomi-Alli, E. O. Omidiora, S. O. Obabiyisi, and J. A. Ojo, "Enhanced E-Banking System with Match-On-Card Fingerprint Authentication and Multi-Account ATM Card," The Journal of Computer Science and Its Applications, International Journal of Nigerian Computer Society (NCS), 19(2), pp. 14-22, 2012.

[24] O. Selina and O. Jane, "Enhanced ATM Security System using Biometrics," (IJCSI) International Journal of Computer Science Issues, Vol. 9, Issue 5(3), 2012.

[25] P. Krishnamurthy and M. M. Redddy, "Implementation of ATM Security by Using Fingerprint Recognition and GSM," National Conference on Research Trends in Computer Science and Technology, International Journal of Electronics Communication and Computer Engineering, Vol.3, Issue (1), 2012.

[26] R. G. Jimoh and A. N. Babatunde, "Enhanced Automated Teller Machine Using Short Message Service Authentication Verification," African Journal of Computing & ICT, IEEE, 7(1), 2014.

[27] K. A. Charles and I. U. Wilfred, "Design of a Secure Unified e-Payment System in Nigeria: A Case Study," African Journal of Business Management Vol. 4(9), pp. 1753-1760, 2010.

[28] D. Maltoni, D. Maio, A. K. Jain, amd S. Prabhakar, "Handbook of Fingerprint Recognition," 2nd edition, Springer, London, UK.

[29] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Second Edition, Elsevier Inc., Rajkamal Electric Press, 2006.

[30] Z. Yang, L. Ying and Y. Gongping, "An Improved K-Means Based Method for Fingerprint Segmentation with Sensor Interoperability," School of Computer Science and Technology, Shandong University, Jinan, China, 2011.

[31] D. J. Fan, B. Sun and J. F. Feng, "Adaptive Fingerprint Segmentation Based on Variance and Its Gradient," Journal of Computer-Aided Design & Computer Graphics, 20 pp. 742-747, 2008.

Universal Telecommunication Access: An Enabling Platform for Implementation of *mHealth* in Developing Economies

Nasir Faruk[1], Olayiwola W. Bello[2], Abdulkareem A. Oloyode[1], Segun I. Popoola[3], Lukman A. Olawoyin[1]
[1]Department of Telecommunication Science, University of Ilorin, Ilorin, Nigeria
[2]Department of Information and Communication Science, University of Ilorin, Ilorin, Nigeria
[3] Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria
Email: faruk.n, laibello, oloyede.aa, olawoyin.la{@unilorin.edu.ng}, segun.popoola@stu.cu.edu.ng

Abstract— **The recent growth in the adoption of mobile health (mHealth) initiatives is mainly driven by low physician-to-patients ratio and prevalent access barriers to remote areas. Mobile technology is increasingly being integrated into health care system for disease surveillance, appointment reminders, disease outbreak alerts, health education and remote access to patient records. In developing countries, these initiatives have the potential to transform the current face of our health delivery service as most emerging economies have long maintained low performance on human development index, doctor-to-patient ratio, and income level. However, successful implementation of these mobile technology-driven initiatives requires low-cost platforms, and more reliable and scalable critical infrastructure deployment beyond legacy ecosystems. Therefore, this paper identified wide gap in digital divide as one of the major implementation challenge of mHealth in developing countries, especially in remote rural areas. Based on these, we propose a universal telecommunication framework that bridges the digital divide to promote deeper penetration of mHealth initiatives, thereby ensuring healthy lives and well-being of all at all ages in line with the 2030 Sustainable Development Goals (SDGs).**

***Index Terms*- rural telecommunication access; software defined radio; spectrum management; business models; regulatory frameworks.**

I. INTRODUCTION

MOBILE health (mHealth) is a subdivision of e-Health (Electronic Health) wherein mobile devices are used to achieve health objectives. Disease surveillance and patient monitoring are considered necessary for effective prevention and control of ill-health. These provide useful information needed for early detection of epidemic and tactical action plan development that will effectively curtail the outbreak. mHealth initiatives have been widely deployed for appointment and medical reminders, interactive patient questionnaire, disease outbreak alerts, health education, remote access to patient registration document, work plan, counselling clinical decision making, and care coordination performance tracking [1]. The International Health Regulations (IHR) has mandated all countries to develop effective disease surveillance and alert systems for proper health response [2].

Globally, there is yet to be any agreed common interoperability standard set for the implementation of mHealth. Nevertheless, it is obvious that voice, text, and data services of mobile communication technologies would be utilized together with underlying technologies such as Global positioning systems (GPS), Wi-Fi, Bluetooth, and other local and personal area network technologies [1]. As of 2011, the global mHealth projects designed for disease surveillance and awareness-raising stood at 26% and 23% respectively [1]. Rapid advancement in mobile technologies and significant increase in mobile subscription and penetration provide good opportunities and platforms for integrating mobile health into existing e-Health services. It is therefore of no doubt that mHealth has the potential to transform the current face of the health delivery service in developing countries.

However, to achieve these goals, mobile network providers has a vital role to play as mHealth platforms require reliable network connectivity (in terms of coverage and capacity) and availability of widely adopted standards and services such as short messaging services (SMS). Successful implementation of mHealth initiatives likewise demands easy and flexible integration of the mHealth network gateway onto the mobile networks. Globally, research efforts have shown that most of the telecommunication infrastructures are deployed in the urban and suburban areas, neglecting the rural areas with limited or no network service [3-5]. Similarly, there exist a wide digital divide between developed and developing countries with respect to ease of access to digital information.

In view of this, the concept of universal telecommunication access is aimed at making telecommunication infrastructure close enough to everyone irrespective of their geographical location, income level, age, gender or other discriminatory parameters. The concept of closeness is defined by availability, affordability and reliability [3]. The major challenges hindering the deployment of rural access schemes includes: population distribution, infrastructure deployment cost, and financial sustainability of rural telecommunication access schemes.

These challenges make it very difficult for mobile operators to deploy macro base stations for cellular services in the rural areas.

The future of the universal access and the benefits of mHealth in developing countries can only be realised if the critical infrastructure and service are presented as low-cost platforms or services that can easily be deployed, beyond the legacy ecosystems. These implementations could leverage ongoing developments in spectrum sharing techniques and lightweight software-based network implementations. This also demands innovations in the regulatory frameworks to create an operating environment that is conducive for innovative or disruptive business models. Therefore, we developed a process framework for universal telecommunication access in developing economies. The outcome of this work will inform policy makers, regulators, and business actors on the potentials of low-cost networks based on spectrum sharing and software defined radio technologies.

The rest of the paper is organized as follows: Section II provides the mHealth architecture and implementation challenges; the level of mHealth adoption in developing economies was unravelled in Section III; Section IV highlights telecommunication access development in developing economies; Section V summarises universal access models (UAM); end-to-end technology framework, and concept for UAM-mHealth are provided in Section VI; and finally, Section VII concludes the paper.

II. mHEALTH ARCHITECTURE AND IMPLEMENTATION CHALLENGES IN DEVELOPING ECONOMIES

A. Architecture

The primary goal of mHealth is to provide medical care and public health using the mobile devices. To fulfil this aim, mHealth architecture must have the capabilities of providing end-to-end medical services to the end users (patients). Mobile communication network is the main driver for successful implementation of mHealth; it serves as the delivery channel as well the interface between the clinicians, patients and electronic records. Fig. 1 shows the bidirectional interactive healthcare services, with direct clinical involvement [6]. The architecture consists of three main segments namely: the mobile network cloud which provides the global connectivity; the clinical services such as the electronic health records (EHR); personal health records (PHR) provided by the clinicians; and the medical application which provides the interface linking the patient mobile devices and the clinicians.

B. Implementation Challenges

Mobile network providers have a vital role to play as mHealth platforms require reliable network connectivity (in terms of coverage and capacity) and availability of widely adopted standards and services such as short messaging services (SMS). In addition to the SMS, some other basic services expected to be provided by the NMO includes; connectivity (voice and mobile broadband), data (HER, PHR and images) and Machine-to-Machine (M2M) services. It is worthy to note that these services will vary, depending on both the network of mobile operators and the capabilities of the mobile phones of end users. Although the current deployed mobile network architecture in most of the developing countries would still support all these services ranging from the SMS reminders to call centers, M2M communication which is considered as future or next generation mobile technology may not be realized at the moment.

III. mHEALTH ADOPTION IN DEVELOPING ECONOMIES

Recently, a couple of mHealth initiatives were unveiled in some developing countries. These include: Mobile Technology for Community Health (MOTECH) in Ghana [7]; mobile phone short message service on antiretroviral treatment in Kenya [8]; AIDS Patient Care using Mobile phones in Uganda [9]; health workers text-message reminders to malaria treatment in Kenya [10]; open-source short message service-based tool for monitoring malaria in remote areas of Uganda [11]; rural health centres, communities and malaria case detection mobile systems in Zambia [12]; cell phone-based and internet-based monitoring and evaluation of antiretroviral treatment in Rwanda [13]; SMS appointment reminders messages in an antiretroviral treatment clinic, in South Africa [14]; mobile direct observation treatment for tuberculosis patients in Kenya [15]; SMS for life in Tanzania [16]; remote clinics with laboratory results via short message service (SMS) in Swaziland [17]; mobile learning system in Botswana [18]; and many other projects reported.

Availability and access to quality healthcare services has been a key challenge in most developing economies. In fact, nearly half of WHO member states (44%) have less than 1 doctor to 1000 patients [19], which is less than the recommended 1:600. Most of these member states are from the developing countries. For example, the ratio of doctor-to-patients in Malawi and Tanzania is 1:50,000, Liberia and Mozambique 1:35,000 [20]. Therefore, the mHealth project initiatives can help to provide better healthcare services in these locations especially, in the rural/remote locations that are not easily accessible. Despite the widening digital divide in developing countries, the adoption of mHealth is critical and important because mobile phone is still more accessible in rural and suburban areas when compared with other basic infrastructures such as electricity and roads networks [1]. The main barriers to adopting mHealth are cost effectiveness, legal issues, knowledge about the system and other competing health proprieties [1]. Still, widespread availability and the recent advances in mobile communication network coverage, and access to cheaper and more powerful mobile devices, provide a good opportunity towards the enhancement, management, prevention and control of chronic diseases.

IV. TELECOMMUNICATION ACCESS DEVELOPMENT AND DIGITAL DIVIDE IN DEVELOPING ECONOMIES

Developing countries contribute largely to the current global high penetration of mobile cellular systems. Whereas, the future of the mHealth in developing countries can only be realised if the required network infrastructure and services are provided, especially in the rural areas. Despite the potential benefits of mHealth in addressing access challenges in rural areas, significant research has revealed the existence of '*digital divide*'. The gap between developed and developing countries, with respect to information availability, continues to increase every day. This can be attested from a wide gap between telecommunication subscription and penetration, "*developed vs developing*". For instance, Fig 2 shows that developing countries have higher subscription in mobile cellular technology than the developed countries; but in reality, the penetration is very low due to the fact that most subscriptions are concentrated in the urban areas. Recent statistics from International Telecommunications Union (ITU) [21] have shown that Africa is the only region where mobile broadband penetration remains below 20%. This could be attributed to the fact that over 60% of African population reside in the rural communities characterized by poor infrastructure, low income, adversely scattered buildings, low literacy level, etc. The mobile broadband penetration in the developing countries stood at 39.1% compared with 82.2% in the developed countries.

Findings from recent exploratory field survey [22] conducted in fifteen rural communities of Kwara State, Nigeria, proved these assertions. Results of the study indicated that out of the 15 villages visited almost all of them do not have access to telephone service. These villages were all less than 40 km from the capital city, Ilorin. The challenges identified remain the population distribution of the rural areas, security and socio-political challenges, infrastructure deployment cost, and financial sustainability of rural telecommunication access schemes. Other works conducted by various international organizations [3-5], such studies cutting across different parts of the world attest this fact.

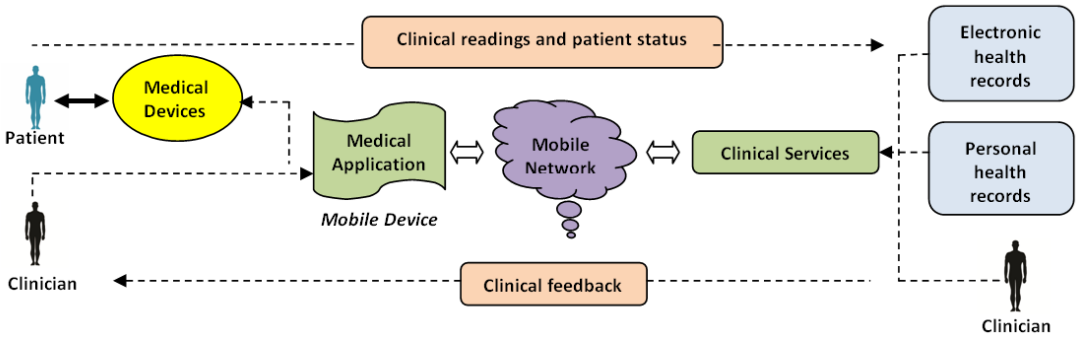
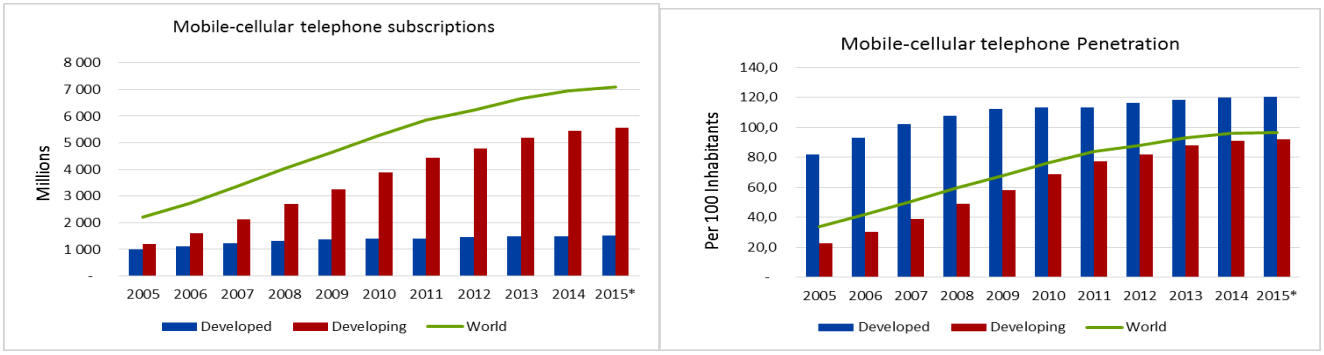


Fig 1: End-to-end Bidirectional interactive healthcare services



(a) (b)

Fig 2. Mobile-Cellular telephone (a) Subscription and (b) penetration (Data source: ITU World Telecommunication/ICT Indicators database).

V. UNIVERSAL TELECOMMUNICATIONS ACCESS MODELS (UAM) FRAMEWORK

The key issue to be considered when developing universal telecommunications access (UTA), particularly for rural communities, is the choice of appropriate approach to achieve the access efficiently and effectively. The approaches, referred to here as universal access models (UAM), depend on many factors which involve the interplay of all the stakeholders: the service providers, governments at all levels, and the benefitting communities. Among the factors to be considered with respect to each stakeholder group, when choosing a UAM, are: the service providers provide level ground for competition, commercial profitability and the tariff packages. The government will provide the capacity to organize and run a fair and open competition, regulations, and monitoring, funding options and identify the actual needs of the communities. The benefitting communities on the other hand are the end-users. Fig 3 shows the schematic diagram of the universal access models framework.

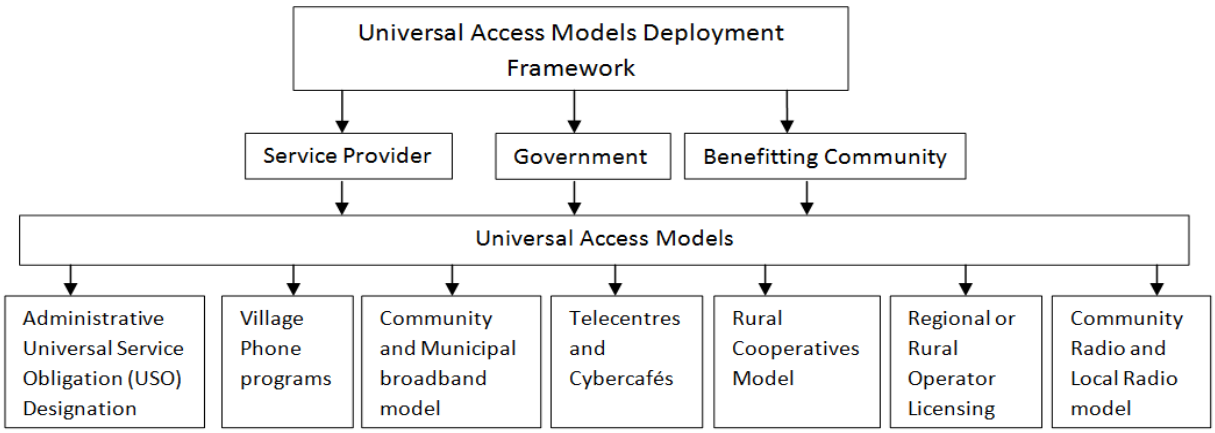


Fig 3: Universal Access Model Framework

VI. END-TO-END TECHNOLOGY FRAMEWORK AND CONCEPT FOR UAM-MHEALTH

A. End-to-End Technology Framework

The future of the universal access and benefits of mHealth in developing countries can only be realised if the critical infrastructure and service are presented as low-cost platforms or services that are easily adaptable, affordable, deployable, configurable and usable, beyond the legacy ecosystems. These implementations could leverage on-going developments in spectrum sharing techniques and lightweight software-based network implementations. The core of this model is the development and promotion of energy efficient and cost effective access networks based on software defined radios and spectrum sharing. The promotion of energy efficiency will directly impact on slowing down the effect of climate change and network energy cost associated with the existing macro radio access networks. This will also free-up power for other development sectors of the economy, especially with the current state of electricity supply in the most developing countries.

Fig 4 depicts the end-to-end technology framework UAM-mHealth. This covers both the backhaul technologies and spectrum management solutions deployed end-to-end between the public Internet domain and the radio access network. The UAM-mHealth framework takes an approach of leveraging existing software defined radio (SDR) platforms as the radio access network. For cost effectiveness, we propose using the Television White spaces (TVWS), long range Wi-Fi or satellite to backhaul the SDR RANs to the core network. In Fig 5, we provide the overall UAM-mHealth framework. In the framework, new players called the Village Service Providers (VSP) are introduced. These VSPs will provide GSM service to various villages and underserved areas using the UAM network access model provided in Fig 3. Disruptive regulatory frameworks and new business models will provide a general framework and ecosystem between the VSPs and the conventional Network service providers (NSP) and backhaul service providers (BSP). The VSP would in turn deploy one or hybrid of the Universal Access Models (UAM) for sustainability and business prospects of their networks within the communities.

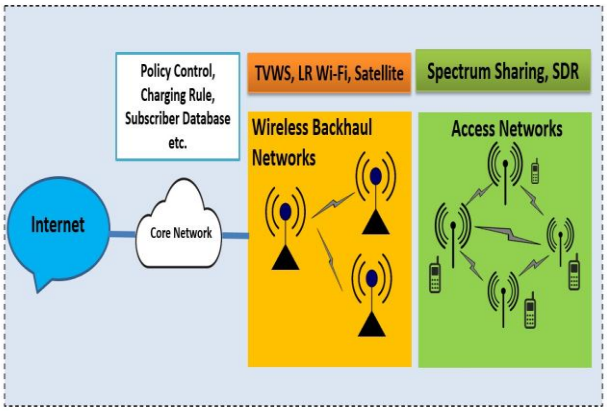


Fig. 4: End-to-End Technology Framework

B. *VSP Network Models Implementation*

The Village Service Providers (VSP) can implement any of the following models:

- *Standalone GSM network Type 1*: A VSP-GSM network consisting of the radio access network, base station subsystem, and direct spectrum assignment (DSA) from the regulators. The backhaul services (BS) are obtained from the NSP or BSP or from both service providers to complement the capacity demand for the VSP. *(See Fig 6).*
- *Standalone GSM network Type 2*: A VSP-GSM network consisting of the radio access network, base station subsystem, and spectrum assignment (i.e. spectrum service (SS)) from the existing mobile network service provider (NSP).
- *Integrated GSM Network:* A VSP-GSM network but integrated with the core network of the existing mobile network operator. The VSP-GSM network will potentially reuse the spectrum of the MNO and the MNO will also provide the backhaul service. The backhaul service could also be sorted from the BSP.

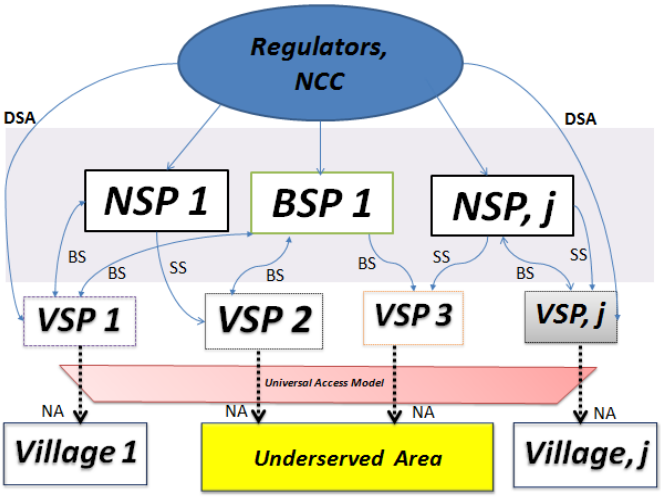


Fig. 5: UAM-mHealth spectrum sharing frame work and backhaul service

VII. CONCLUSION

Mobile Health has been identified as the easiest way of extending healthcare delivery service and disease surveillance in developing economies, particularly in remote areas with difficult terrains. The rapid advancement in the mobile technologies, increased in mobile subscription and penetration, provide opportunities and enabling platform for integrating mobile health into existing e-Health services. Findings from this paper show that considerable mHealth initiatives have been proposed in many developing countries. It is also found that, the future of the mHealth in developing countries can only be realised if required reliable network infrastructure and affordable services are provided, especially in the rural areas. Digital divide is a principal limiting factor affecting wider coverage, adaptation, and implementation of mHealth initiatives. The gap between developed and developing countries, with respect to access to telecommunication service is becoming increasingly wide. For effective implementation of mHealth in developing countries, the mobile network providers must provide at the minimum, the necessary connectivity (coverage and capacity) and facilitate the use of the widely adopted standards and services such as SMS. The proposed framework would help in bridging the digital divide and in extending mHealth initiatives in the rural areas towards achieving the 2030 Sustainable Development Goals (SDGs).

REFERENCES

[1] WHO, "New horizons for health through mobile technologies: Global Observatory for eHealth series", WHO Library Cataloguing-in-Publication Data, ISBN 978 92 4 156425 0, vol 3, pp 2-112, 2011.

[2] http://www.who.int/topics/international_health_regulations/en/ visited 19/08/2016

[3] Xavier P. "Universal *Access for telecommunications in a competitive environment".* OECD-WORLD BANK Fifth Services Expert Meeting, 3-4 February, 2006.

[4] OECD, "*Universal Service Policies in the Context of National Broadband Plans"* Working Party on Communication Infrastructures and Services Policy, Directorate for Science, Technology and Industry Committee for Information, Computer and Communication Policy, Organisation for Economic Co-operation and Development.

[5] Kawasumi Y. "*Rural Communications on the Global Agenda"*, Global Survey on Rural Communications, International Telecommunicaion Union, http://www.itu.int/itunews/manager/display.asp?lang=en&year=2004&issue=05&ipage=globalAgenda

[6] GSMA,"A High Level Reference Architecture for Mobile Health", GSMA, vol. 1, November, pp 1-40,

[7] Grameen Foundation. Mobile Technology for Community Health (MOTECH) project in Ghana, Early lessons learned; March 2011.

[8] Ngugi E, Ball TB, Thabane L, Gelmon LJ, Kimani J, Ackers M, Plummer FA: Effects of mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010, 376:1838–1845.

[9] Chang LW, Kagaayi J, Nakigozi G, Packer AH, Serwadda D, Quinn TC, Gray RH, Bollinger RC, Reynolds SJ: Responding to the human resource crisis: peer health workers, mobile phones, and HIV care in Rakai, Uganda. AIDS Patient Care STDS 2008, 22(3):173–174.

[10] Zurovac D, Sudoi RK, Akhwale S, Ndiritu M, Hamer DH, Rowe AK, Snow RW: The effect of mobile phone text-message reminders on

Kenyan health workers' adherence to malaria treatment guidelines: a cluster randomised trial. Lancet 2011, 378:795–803.

[11] Asiimwe C, Gelvin D, Lee E, Amor YB, Quinto E, Katureebe C, Sundaram L, Bell D, Berg M: Use of an innovative, affordable, and open-source short message service-based tool to monitor malaria in remote areas of Uganda. Am J Trop Med Hyg 2011, 85(1):26–33.

[12] Kamanga A, Moono P, Stresman G, Mharakurwa S, Shiff C: Rural health centres, communities and malaria case detection in Zambia using mobile telephones: a means to detect potential reservoirs of infection in unstable transmission conditions. Malar J 2010, 9:96.

[13] Nsanzimana S, Ruton H, Lowrance D, Cishahayo S, Nyemazi J, Muhayimpundu R, Karema C, Raghnathan P, Binagwaho A, Riedel D: Cell phone-based and internet-based monitoring and evaluation of the national antiretroviral treatment program during rapid scale-up in Rwanda: TRACnet, 2004-2010. J Acquir Immune Defic Syndr 2012, 59(2):e17–e23.

[14] Crankshaw T, Corless IB, Giddy J, Nicholas PK, Eichbaum Q, Butler LM: Exploring the patterns of use and the feasibility of using cellular phones for clinic appointment reminders and adherence messages in an antiretroviral treatment clinic, Durban, South Africa. AIDS Patient Care STDS 2010, 24(11):729–734.

[15] Hoffman JA, Cunningham JR, Suleh AJ, Sundsmo A, Dekker D, Vago F, Munly K, Igonya EK, Hunt-Glassman J: Mobile direct observation treatment for tuberculosis patients: a technical feasibility pilot using mobile phones in Nairobi, Kenya. Am J Prev Med 2010, 39(1):78–80.

[16] Barrington J, Wereko-Brobby O, Ward P, Mwafongo W, Kungulwe S: SMS for Life: a pilot project to improve anti-malarial drug supply management in rural Tanzania using standard technology. Malar J 2010, 9:298.

[17] Jian WS, Hsu MH, Sukati H, Syed-Abdul S, Scholl J, Dube N, Hsu CK, Wu TJ, Lin V, Chi T: LabPush: a pilot study of providing remote clinics with laboratory results via short message service (SMS) in Swaziland, Africa. PLoS One 2012, 7(9):e44462.

[18] Chang AY, Ghose S, Littman-Quinn R, Anolik RB, Kyer A, Mazhani L, Seymour AK, Kovarik CL: Use of mobile learning by resident physicians in Botswana. Telemed J E Health 2012, 18(1):11–13.

[19] WHO, "Global Health Observatory (GHO) Data". datahttp://www.who.int/gho/health_workforce/physicians_density/en/

[20] ----------"Worldwide Doctor-Patient ratio", http://www.gmu.ac.ae/careandshare/worldwide.html visited 08/09/2016

[21] ITU," ICT facts and Figures" https://www.itu.int/en/ITUD/Statistics/Documents/facts/ICTFactsFigures2015.pdf visited 10/07/2016.

[22] Adediran Y.A, J.F. Opadiji, , N. Faruk, O. W. Bello,"*Rural Telecommunications Access in Nigeria*, Final Report of 2014/2015 TETFund Research Grant, University of Ilorin, Ilorin, Nigeria, pp 1-180, 2015.

Review on Data Mining Methods for Prediction of Diseases in Healthcare

Ogundokun, Roseline Oluwaseun
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
seunlayan@gmail.com

Abikoye, Oluwakemi Christiana
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
Kemi_adeoye@yahoo.com

Aro, Taye Oladele
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
taiwo_aro@yahoo.com

Adegbola, Ismail A.
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
ismailadegbolaa@gmail.com

***Abstract*— In several organizations and industries especially in healthcare, data mining has been of great importance and is gaining more attention in the technological society. Data mining is employed in healthcare for managing patient information, disease diagnosis, disease prediction and fraud detection in the health insurance services. This paper discusses different techniques use in data mining to diagnose, manage and predict intimidating diseases such as heart diseases, cancer, lungs diseases, kidney diseases, diabetes, and tuberculosis.**

Keywords—— Data mining, Data mining techniques, Diagnose, Healthcare, Prediction

I. INTRODUCTION

The healthcare system is a system that involves the proper care or improvement of individual health conditions through several processes such as diagnosis, treatment, management and prevention of diseases and other physical and mental impairments. Healthcare services are rendered by health practitioners in collaboration with many health workers, chiropractic, physicians, dentistry, midwifery pharmacy and optometry. A main problem contending healthcare systems is the delivery of standard service at reasonable costs. For a sick person to be offered a quality and effective treatment by medical personnel it requires correct diagnosis. Data mining technique in health care a process that involves analyzing of data from different perspectives and reducing it into more relevant information that can be used for prediction and diagnosis of some diseases [1].

Techniques in data mining are process that involve the extraction and discovery of hidden patterns from large raw datasets [2]. The data mining technique can be used to formulate knowledge out of data then representing it in a manner that is easily understood by people or persons. The applications of data mining in healthcare system include analysis of healthcare centers for quality policy making and treatment faults avoidance, on time detection, prevention of diseases, inevitable hospital deaths and identification of fake insurance claims. [3].

Several applications can benefit by the use of information and knowledge obtained from huge amount of data. Knowledge discovery is seen as a process of non-trivial extraction of information from a large dataset [4]. The information which is previously concealed in database is potentially useful for further discovery of facts.

II. KNOWLEDGE DISCOVERY IN DATABASE (KDD)

KDD procedure is a method of computation that finds out patterns in large datasets as clearly shown in figure 1. KDD involves the selection of relevant data, processing it and converting the data into useful information for extraction of hidden information. The KDD process can be further categorized into:-.

A. Selection

The procedure of obtaining relevant data for the task of analysis from the database. The focus here is on a subset of data variables or samples on which discovery is to be performed.

B. Preprocessing

It is a very important step in KDD, if data is not adequately screened this may affect the final results [1]. This process is necessary for data cleaning. It removes noise and inconsistent data and combines multiple data sources. The strategies are designed for handling missing data fields.

C. Transformation

Converts data into appropriate patterns before performing techniques in data mining. It reduces the effective variables into a considerable amount and also locates invariant representations for the data.

D. Data integration

Combines data from different locations into a comprehensible database as the collection of data are always from diverse data stores. Microsoft excel is usually used to merge data together in data mining.

E. Interpretation / Evaluation

Interprets patterns into information by the removal of redundant, inappropriate data and converting the useful patterns into terms that are understandable.

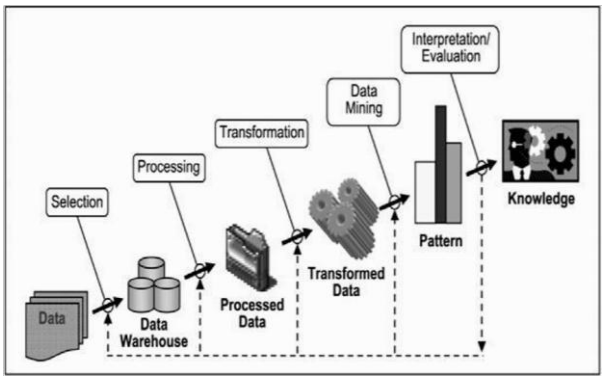


Fig.1. Procedure in Data mining method [5]

III. ABBREVIATIONS AND ACRONYMS

KNN	K-Nearest Neighbour
UCI	UC Irvine health
SVN	Support Vector Machine
FP Tree	Frequent Pattern Tree
Kstar	Is an instance based classifier
WEKA	Waikato Environment for Knowledge Analysis
CRISP-DM	Cross Industry Standard Process for Data Mining
SQL	Style Query Language
KDD	Knowledge Discovery in Databases
DMX	Data Mining Extension
WAC	Weighted Association Classifier
ANN	Artificial Neural Network

IV. RELATED WORK

Researchers have conducted several studies in relation to diseases prediction and diagnosis using different data mining techniques. They made contributory efforts in providing efficient methods of data mining technique in term of accuracy in finding prediction for some diseases that are major concerned to humans. The following studies have employed data mining methods for prediction of numerous diseases:

[6] presented a study on techniques of data mining for prediction of heart disease. Risk factors for heart disease were also mentioned. They analyzed various techniques in data mining for prediction of heart disease. Performances study of Naïve Bayes, Decision tree, KNN was also analyzed.

Comparisons were conducted between Neural Network, Fuzzy Logic and Decision tree. Heart disease classification system was also reviewed and from the result, it was established that techniques in data mining play major roles in classification of heart diseases. It was also deduced that Neural Network is early stage performed very well for prediction of disease. Different technologies used showed different accuracies depending on the number of attributes considered and data mining tools used for implementation.

[1] proposed a study to compare the accuracy in performance of Multilayer perceptron Neural Network, Iterative Dichotomiser 3, C4.5 and Decision tree algorithms for typhoid fever disease prediction. The data used was gathered from a well-known Nigerian hospital. The collected data was converted to acceptable form in data mining. In order to perform prediction, the training set was enabled to allow system observe relationships between input and output data. The database used for testing contained data used to test performance of the model MLP ANN gave a better performance due to low values generated in the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Relative Absolute Error (RAE) error performance measures.

[7] used Kernel Discriminant Analysis (KDA) and Neural Network techniques in data mining to forecast diabetes mellitus. They used database that contained 768 people data collected from UCI dataset with 9 risk factors. The risk factors used included diastolic blood pressure, blood cholesterol, glucose, plasma, insulin, BMI, DPF, Triceps (SFT) and age class. The KDA was used to find a maximum transformation between the class variance and minimizing within the class variance. The feature of risk factors obtained using KDA were supplied as input by the feed forward system, on the basis these

features, the feed forward produced signals as output. These signals were output that showed either a person has diabetes. The result indicated that Neural Network and KDA method outperformed the convectional Artificial Neural Network in average prediction accuracy.

[8] employed approach of two layers for identification of heart disease possibility. The first layer contained a very high factors that are compulsory for occurrence of coronary heart disease and the second layer is the remaining factors taken at the second level. The 2 levels approach increased the performance of their system as it aided the prediction of heart disease chances precisely. The disease of heart database was obtained from UCI machine learning database which was used to train the neural network and then applied fuzzy rules for prediction of chances of coronary heart diseases as low, average or high. Finally, the two levels method were considered to get a final decision. The applied neurofuzzy integrated approach at two levels gave reduced error rate and high work efficiency.

[9] proposed model to predict kidney disease by using Artificial Neural Network (ANN) and Support Vector Machine (SVM). A comparative analysis was conducted for performance of the two algorithms on the bases of it accuracy and execution time. The prediction of four types of diseases of kidney (Acute Nephritic Syndrome, Chronic Kidney disease, Acute Renal Failure and Chronic Glomerulonephritis was the major focused of their study. From the experimental result, it was reviewed ANN outperformed SVM.

[10] applied K-nearest neighbor (KNN) to aid the medical experts in the prediction of disease of heart. They considered in their work that if incorporating voting with KNN would improve disease of heart prediction accuracy. The result reviewed that using KNN would offer better accuracy than neural network ensemble in the prediction of heart disease. Also the results obtained showed the application of voting had no significant improvement on KNN accuracy in the heart disease prediction. They concluded that KNN algorithm proved to be more accurate when applied with fuzzy rules not for all disease but for particular set of disease.

[11] reviewed several literatures on application of different methods of data mining for the heart ailment prediction. The main objective of their work was to study different approaches in data mining which are used in disease of heart prediction. The study also aimed at analyzing different data mining techniques that has been introduced in recent years for heart disease prediction system. They concluded that employing data mining in prediction of heart disease will further assist the medical experts to diagnose in less time and also give a prediction of future complications

[12] evaluated performance of Neural Network application for prediction of heart disease. A predictive system for heart disease was developed using Artificial Neural Network and Genetic Algorithm technique. The system calculated the number of hidden nodes for Neural Network which trained the network with proper selection of neural network architecture and applies the global optimization method of genetics algorithm to initialize Neural Network. The proposed system trained the network by employing back propagation algorithm method using weights optimized by genetic algorithm. The genetic neural approach for disease of heart prediction was used to test data to the optimum value and predict whether a patient is having heart disease. Data for risk factors in connection to disease of heart were obtained from 50 persons provided by American heart Association. The result showed that predictive system with genetic neural method gave 98% accuracy.

[13] presented a prototype system for the prediction of disease of the heart and cancer of the breast using methods of data mining. They used publicly available data from web containing 909 records for disease of the heart and 699 for cancer of the breast. The C4.5 and the C5.0 decision tree algorithms were used on the database for prediction and their comparative performance analysis were carried out. The study also presented how these rules can be used in proof established medicine.

[14] developed a system for liver disorder analysis using data mining approaches. They got their data from UCI database which contained about 345 instances with 7 unlike attributes. The instances were pertaining to two types of blood test which are considered as variant to disorders in the liver that might arise from high volume of alcohol consumption. WEKA Tool was used for the classification of the data and the data were also evaluated using 10-fold cross validation and the results were compared. WEKA tool was used to compare the performance accuracy of data mining algorithms for diagnosis of liver disease datasets. The feature selection in the tool describe the attributes status of the data present in the liver disease. They used three supervised machine learning algorithms; Naïve Bayes, FP trees and Kstar. The three algorithms were used to predict the accuracy of liver disease. FP tree algorithm performed better because it only took lesser time to calculate the accuracy than other algorithms.

[15] presented performance analysis on classification methods in data mining over heart disease database. They used Naïve Bayes and WAC (Weighted Association Classifier) data mining classification modelling methods due to their ability to answer difficult questions which are found not easy with traditional decision support system. The system obtained concealed knowledge from old heart disease database. Data Mining Extension (DME), a query

language and functions were applied to develop and access the system. Classification matrix methods were used to evaluate performance and effectiveness of the model.

[16] applied K-Nearest Neighbour (KNN) for diagnosis of heart disease. They investigated whether the integration of voting with KNN can improve accuracy for diagnosis of heart disease. The experimental result showed that using KNN achieved 97.4% accuracy. The acquired accuracy is higher than any other published findings on the benchmark database. The study again reviewed that applying KNN could reach greater accuracy than Neural Network ensemble in the diagnosis of heart disease. They mentioned that application of voting could not improve the accuracy of KNN in the diagnosis of heart disease.

[17] proposed a system for diagnosis of heart disease using decision support system together with techniques in data mining. They employed Naïve Bayesian and K-Nearest Neighbour (KNN) as data mining tools for prediction. The system generated and obtained unknown information (patterns and relationships between different features from the historical heart related database). The developed system answered difficult questions which cannot be answered by traditional decision support system. The system designed can be used for as training system for students in the medical school.

V. CONCLUSION

Data mining techniques are approaches employ for sorting out data to find hidden patterns in order to establish knowledge. These techniques have been used in the healthcare systems to predict and diagnose different types of diseases. This paper discussed a review on several techniques in data mining used over the years for the prediction of diseases in health care industry.

References

[1] O. O. Adeyemo and T. O. Adeyeye "Comparative Study of ID3 / C4. 5 Decision tree and Multilaye Perceptron Algorithms for the Prediction of Typhoid Fever, " African J. Comput ICT, vol. 8, no. 1, pp. 103-112, 2015

[2] A. S. Aslam and I, "Data Mining Algorithms and their applications in Education Data Mining," Int J. Adv. Res. Comput. Sci. Management Stud, vol. 2, no. 7, pp. 50-56, 2014.

[3] R. Jagtap and S. Joshi, "Hybridization of Data Mining Techniques in Heart Disease Diagnosis and Treatement," Int. J. Eng. Res. Technol, vol. 3, no. 12, pp. 596-598, 2014.

[4] T. Smitha and V. Sundaram, "Knowledge Discovery from Real Time Database using Data Mining Technique," Int. J. Sci. Res. Publ., vol. 2, no. 4, pp. 2-4, 2012.

[5] T. Karthikeyan, B. Ragavan, and V. A. Kanimozhi, "A study on Data Mining Classification Algorithms in Heart Disease Prediction, "Int. J. Adv. Res. Comput. Eng. Technol., vol. 5, no 4, pp. 1076-1081, 2016.

[6] D. C. Bindushree, "Prediction of Cardiovascular Risk Analysis and Performance Evaluation Using Various Data Mining Techniques: A Review, "Int. J. Eng. Res., vol. 5013, no. 5, pp. 769-800, 2016.

[7] K. Singh, S. and Kaur, "International Journal of Advanced Research in Computer Science and Software Engineering Diabetes Mellitus Forecast using Kernel Discriminant Analysis (KDA) with Neural Network Classifier," Int. Jornal Adv. Res. Compyu. Sci. Softw. Eng., vol. 5, no. 9, pp. 789-792, 2015.

[8] A. K. Sen, S. B. Patel, and D. P. Shukla, "A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level," Int. J. Eng. Comput. Sci, vol. 2, no. 9, pp. 1663-1671, 2013.

[9] D. S. Vijayarani and S. Dhayanad, "Kidney Disease Prediction Using SVM and ANN," Int. J. Comput. Bus. Res., vol. 6, no. 2, 2015.

[10] S. P. Chaudhari, A. A. & Akarte, "Fuzzy & Data Mining based Disease Prediction Using K-NN Algorithm," Int. J. Innov. Eng. Technol. Vol. 3, no. 4, pp. 9-14, 2014.

[11] N. Kundra, Di & Kaur, "Review on Prediction System for Heart using Data Mining Techniques, Int. J. Lastet res. In Engineering Technol., vol. 1, no. 5, pp. 9-14, 2015.

[12] N. P. Waghulde and N. P. Patil, "Genetic Neural Approach for Heart Disease Prediction, "Int. J. Adv. Comput. Res., vol. 4, no. 3, pp. 778-784, 2014.

[13] M. T. Khan, S. Qamar, and L. F. Massin, "A Prototype of Cancer / Heart Disease Prediction Model Using Data Mining," Int. J. Appl. Eng. Res., vol. 7, no. 11, 2012.

[14] P. R. G. S. Reena, "Analysis of Liver Disorder Using Data Mining Algorithm," Glob. J. Comput. Sci. Technol., vol. 10, no. 14, pp. 48-52, 2010.

[15] N. A. Sundar, P. P. Latha, and M. R. Chandra, "Performance Analysis of Classification Data Mining Techniques over Heart Disease Database," Int. J. Eng. Sci. Advanced Technol., vol 2, no. 3, pp. 470-478, 2012.

[16] M. Shouman, T. Turner, and R. Stocker, "Applying k-Nearest Neigbour in Diagnosing Heart Disease Patients," Int. J. Inf. Educ. Technol., vol. 2, no. 3, pp. 220-223, 2012.

[17] P. Shinde, S. B. & Amrit, "Decision Support System on Prediction of Heart Disease Using Data Mining Techniques," Int. J. Eng. Res. Gen. Sci., vol. 3, no. 2, pp. 1453-1458, 2015.

Prototype of Molecular Biology Courseware for Inclusive Education System: An Instructional Interface

Abdulrauf Tosho
College of Physical Sciences
Al-Hikmah University, Ilorin,
Kwara State, Nigeria
abdtosh@gmail.com

Shakirat Haroon Sulyman
Department of Information and Communication Science
University of Ilorin
Kwara State, Nigeria
shakiraharoon@yahoo.com

***Abstract*— This paper provides a research approach in developing an e-learning environment for inclusive learners. By capturing and modeling the multimedia 3-D knowledge into an instructional interface for teaching and assessing Molecular Biology Courseware for student in distance learning. The research proposed and developed a Conceptual Design Model Courseware for Inclusive Education System, called C4IES. The experimental result from the subject experts that used the proposed prototype has shown a great potential in the courseware usage. Where, out of 10 experts, 90% agreed that the proposed Multimedia environment has demonstrated a high usability in enhancing molecular concept (DNA) and practical skills of student in distance learning and also for regular students in practicing the laboratory task while away from on-campus environment. This could also be as guidance to the developer or anyone who intends to develop courseware for an inclusive environment.**

Keywords—courseware; instructional; inclusive learners; molecular biology; strategy; usability.

I. Introduction

Everybody has equal rights to quality learning activities. Technique to ease the learning acquisition needs to design for every user, including both disabled and non-disabled. However, for the inclusive, learning instructional is quite demanding. With the setback facing those with impaired, various challenges in the learning instructional interface are higher, which finally could result to learning difficulty. Those with special needs have their own setback that makes them incapable to learn at the same pace with normal persons [1]. They require specialized interface learning instructions in order to maximize their potential and self-sufficiency learning attributes. Among many types of the disabilities, visual and hearing impairment is considered as non-mental disorder that has ability to learn with non-disabled learners in an inclusive environment. [2] Reports anticipated that there are 15% of the world population has some form of disabilities. Department of Social Welfare has registered the total number of billon disabled people in December 2012 [3]. The report mentioned the total number of disabled people in Malaysia is 305,640. However, these data are incomplete as registration of persons with disabilities in Malaysia is not compulsory, and is done only on a voluntary basis. In addition, the data are not up to date, as the names of those who have died are not deleted from the main record [4].

The facts still remains that the number of disable keeps on increasing drastically. Therefore, exposing them to the inclusive education system is important because they should together be respected as part of the resources in the country. Unfortunately, study reveals that 80% of learning materials such as textbook and courseware are provided for different target learners [5] but are not for inclusive users. This is because the main learning styles that are used for normal learning for disabled students are followed by text reading or kinesthetic [6]. In terms of definition, Inclusive Education is where learners with specific needs or learning disabilities learn alongside learners with no disabilities in the same environment [7]. This is to place suitable special needs learners in tertiary institutions either part-time or full-time, and in-school or distance learning according to their capabilities. The effort of inclusive education system has been undertaken by both developed and developing countries. For instance, Malaysia Education Blueprint (MEB) 3013 – 2025 ensures to have students whose circumstances or needs are learning alongside with non disabled student in the mainstream that are likely to fall through the gaps in education are especially catered for, while this will help them reach their full potential. Presently, there are several problems in the area of creating health community awareness through learning approach. For instance, the interface design is vague or complex for the users, and the content lacks motivational elements to serve usability strategies for creating community health awareness [8]. In the area of instruction, learners have difficulties in using interface design tools to guide their learning [9], and content acquisition methods are often not that effective [10]. Consequently, learning abstractly about molecules regarding 2-D interface is often reported in literatures. This problem related to the poor user interface and

inappropriate structure of instructional elements to present information [11]. As a result, the learning model and means are rigid [12].

This indicates that inclusive learners need typical learning materials that specifically could fulfill their needs in learning without facing any more difficulties, particularly in terms of content acquisition, navigation accessibility, and usability strategy aspects. Instructional interface with multimedia could be a better way that can enhance the understanding and content acquisition in learning materials for an inclusive environment, this occurs when the user can control 'what', 'when' and 'how' of such elements, which includes text, audio, video, graphics and animations [13]. It has the capacity to deliver learning materials in multiple forms which can motivate any form of learners with limiting in the specific learning difficulties.

In response to that, this study attempts to develop a prototype of courseware which specifically caters for the needs of inclusive learners in learning, which is named Courseware for Inclusive Education System (C4IES). Prior to developing the C4IES, a set of specific design principles has to be determined in making sure C4IES could fulfill the needs of both impaired and non-impaired learners. Hence, with the support of critical analysis carried out, this study comes out with two specific objectives as i) to determine the design principles of C4IES, and ii) to develop a prototype of C4IES based on the gathered design principles. Thus, in achieving both objectives, two phases of activities were performed as discussed in the next section.

II. PHASES OF ACTIVITIES AND METHODS

This study involves two phases of activities which are identification of element specification that result of conceptual design for C4IES model, and prototype development through IntView methodology as shown in the Figure 1. In the first phase, the activities involved include literature study and preliminary study. This technique is known as User Centered Design (UCD) approach. From this phase, data regarding the design concepts of C4IES model were gathered and the first objective of the study was achieved.

III. C4IES PROTOTYPE DEVELOPMENT

In the development process of C4IES prototype, **IntView** methodology was adopted. It involves two phases, which are pre-production, and production. In the first phase, 11 steps were implemented. In developing C4IES it is important to involve users and experts before the development of C4IES begins. The development of C4IES prototype takes up the challenge to ensure opportunities of multimedia interactive courseware presentation to cater for content acquisition and accessible to all learners, including those with various impairments. The learning materials for learners of different target were used to gather the input and comments in terms of the design of C4IES prototyping. All this input is important in preparing the script and storyboard of C4IES. Therefore, the subject area of learning content in C4IES is Molecular Biology This is based on the finding from preliminary studies on the subject taken has shown that students have problems in learning biology course. According to [14], students having problem in visualizing the concept of the Molecular Biological Cell in abstract reading, in un-interactive environment and the images are represented in 2-Dimensional in the textbook. Hence, the prototyping of interactive courseware with multimedia elements was developed for this topic.

In the development phase, all the identified elements and design from the previous stage were referred for developing the working prototype. The early step, before commencing the development process is to identify the required tools. In this research, the development tool being chosen is Adobe Photoshop CS6 which plays different roles in the development process of the prototyping. Apart from the technological tools, the development process of "Learning Content" module highly relies on the learning material on the PDF lesson material, which as the main source of content development and design based on the identified instructional strategies that proposed for the C4IES model [15]. Cross reference was also made by referring to several available reference books. This is to make sure that the multimedia courseware prototype is rich in contents and informative for the students. The modules for C4IES are divided into different units which are as shown in Figure 2.

Based on the determined usability strategies for the proposed conceptual model of C4IES, this study comes up with specific instructional elements that serve as usability strategy, design principles that has been enhanced the content accessibility and consider the needs of different abilities in the learning content (inclusive user). Theses make the courseware usable for particular disability groups; and they come in the form of implementing assistive technologies or retrofitting the courseware. Table 1 shows the lists of instructional strategy principle for particular user groups. However, [16] have been supposing that non-impaired people and hearing-impaired people have little difference in interface use, except for the sounds in the contents design. Therefore, the study provides equivalent alternatives to auditory and visual content for hearing-impaired learners. Specifically, this guideline has been stated as the following: "Describe the sound of auditory content" and "Provide non-text equivalents of text" Furthermore, Figures 3 through 6 show some interface samples of the C4IES prototyping snapshots that developed in this study based on the courseware module design and instructional strategy principles as determined in the previous phase.

IV. CONCLUSION AND FUTURE WORKS

Summarily, this study reports an ongoing project regarding the development of C4IES prototype. C4IES was developed based on the identified instructional strategy principles in an attempt to fulfill the needs of inclusive learners in learning activities. Content acquisition, navigation, accessibility, and motivation are applied in C4IES through the identified instructional strategy design principles. Future works of this study are to investigate the user experience of using C4IES in terms of motivation to use courseware again next time,

efficiently in shorting the time to accomplishing a task, and effective for inclusive learning environment.

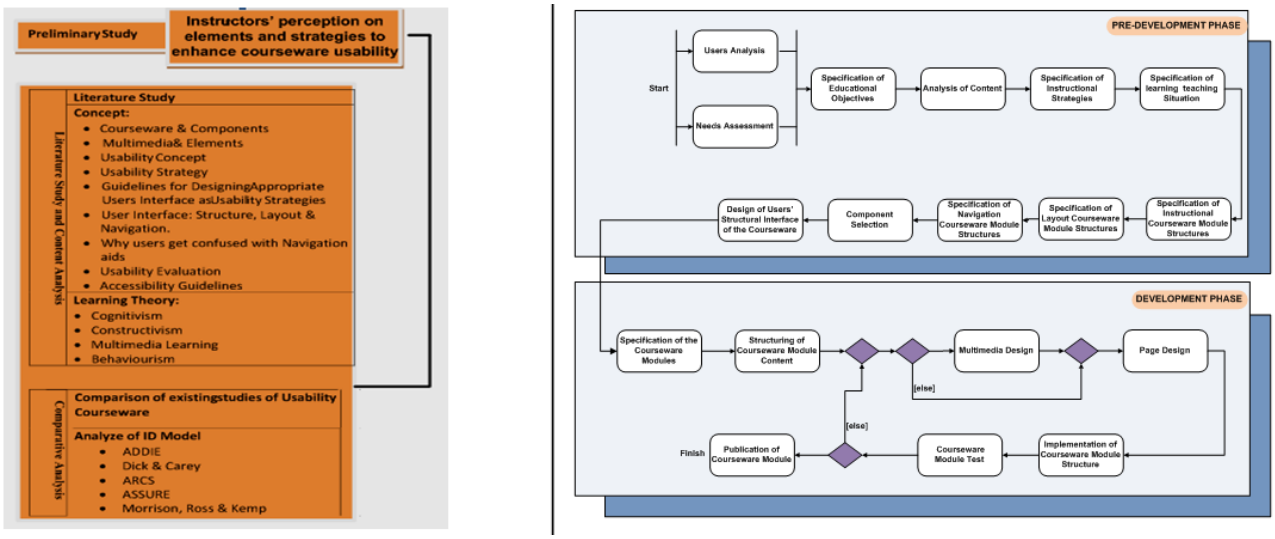


Figure 1: Summary of activities

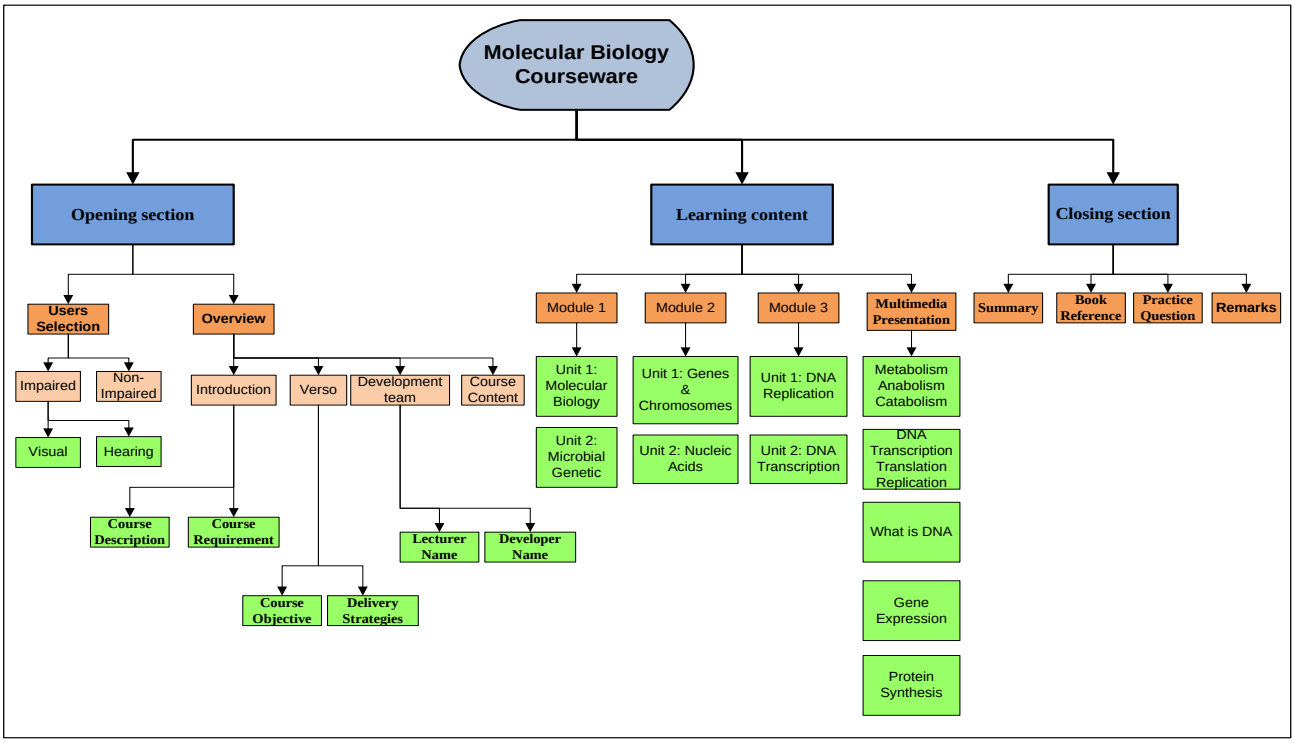


Figure 2: Courseware Modules

TABLE 1: INSTRUCTIONAL ELEMENTS FOR USABILITY STRATEGIES FOR DIFFERENT ABILITIES (DESIGN PRINCIPLES)

Non-impaired	Low Vision impaired	Hearing impaired
✓ Avoiding large graphics. ✓ Avoiding numerous graphics. ✓ Avoiding lengthy pages of content presentation. ✓ Offering navigation support. ✓ Simplifying user interface. ✓ Avoiding complexity design. ✓ making content succinct and relevant. ✓ Using text to label images. ✓ Provide users' control on sounds. ✓ Provide a text equivalent. ✓ Provide feedback. ✓ Provide role-over on navigation button key.	✓ Provide enough contrast between text and background color. ✓ Text -to- speech system. ✓ Provide an auditory equivalent. ✓ Colors Mindful for action Items. ✓ Provide closed captions for all audio content that contains useful information. ✓ Use the largest font size. ✓ Auditory feedback. ✓ Tactile interface. ✓ Closed captions. ✓ Screen/image-enlargement utility. ✓ Provide role-over on button key. ✓ Object magnification. ✓ Avoid blinking, flickering, or moving elements. ✓ Do not design something differently from user expectations just to be different. ✓ Eliminate unnecessary complexity.	✓ Speech-to-text system. ✓ Closed captions. ✓ Provide a text equivalent. ✓ Conform to current interface design standards. ✓ Provide transcripts for all audio content. ✓ Provide captions and descriptions of multimedia used. ✓ Using text to label images. ✓ Turn off graphics. ✓ Turn off sounds. ✓ Provide enough contrast between text and background color. ✓ Use descriptive links rather than "click here. ✓ Use the largest font size. ✓ Provide feedback. ✓ Avoid blinking, flickering, or moving elements. ✓ Do not design something differently from user expectations just to be different. ✓ Eliminate unnecessary complexity.

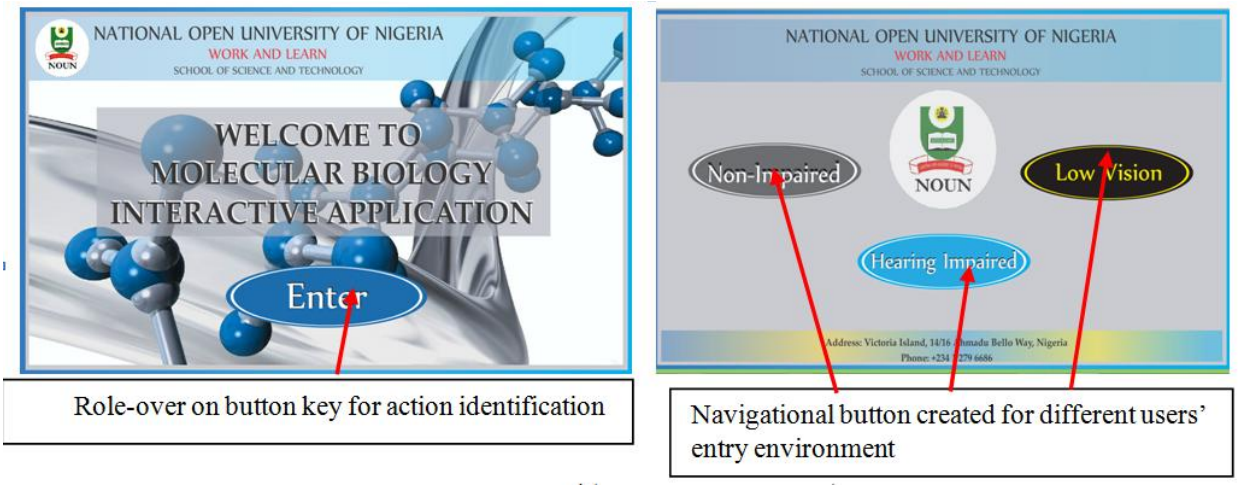
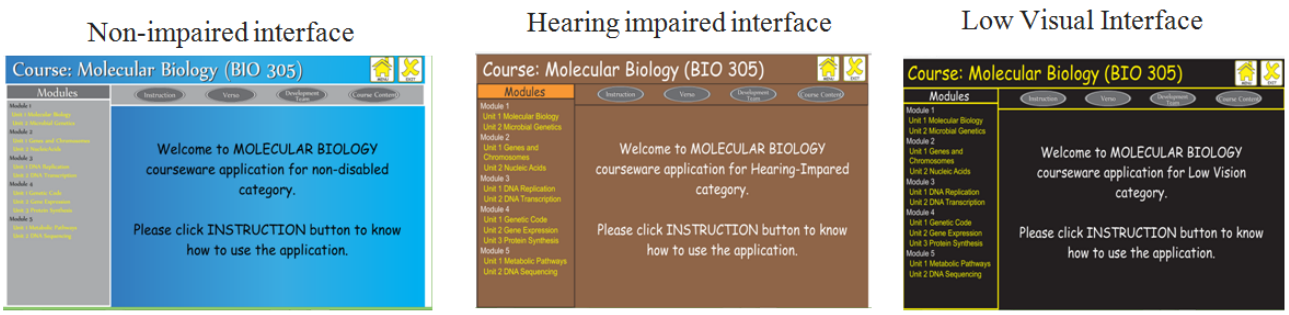


Figure 3: Structural (Opening section in C4IES)



Create good contrast color between background and foreground: As example use either black or deep color for text and bright color for background

Figure 4: Interface design (Layout section in C4IES)

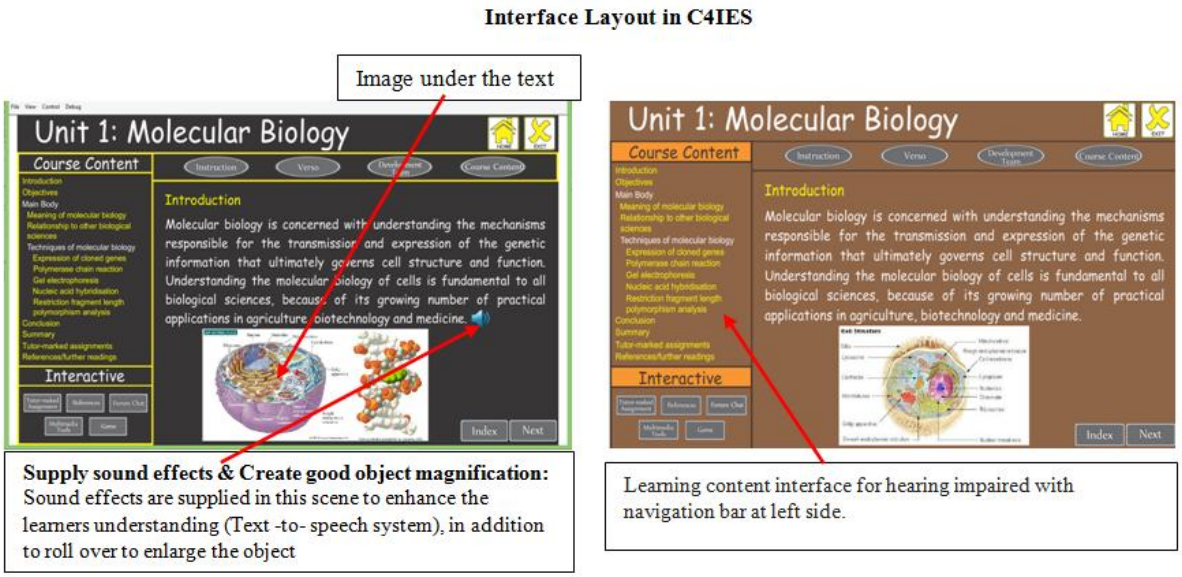


Figure 5: Content presentation in C4IES

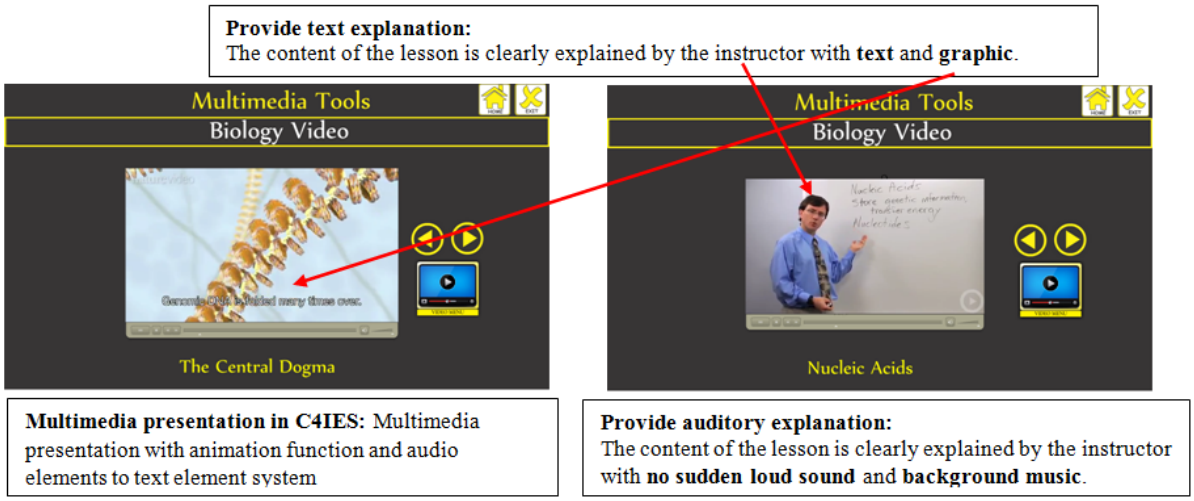


Figure 6: Multimedia interactive instructive section in C4IES

REFERENCE

[1]. Abdollah, N., Ahmad, W. F. W., and Akhir, E. A. P. (2010, May). Multimedia Design and Development in Komputer Saya'Courseware for Slow Learners. In Computer Research and Development, 2010 Second International Conference on (pp. 354-358). IEEE.

[2]. World Health Organization. (2012). World report on Disability.[Online].Available:http://www.who.int/disabilities/world_report/2011/report.pdf. [Accessed:August. 11, 2015].

[3]. Rashid, M. Z. (2010). "Empowering persons with disabilities in Malaysia," 10th BJM of the East Asia Pacific Regional Council of Cheshire Home, pp. 8-11.

[4]. Tiun, L. T., Lee, L. W., and Khoo, S. L. (2011). Employment of people with disabilities in the Northern State of Peninsular Malaysia: Employers' perspective. Disability, CBR and Inclusive Development Journal, 22, 79-94.

[5]. Chadha, R. K., and Subramanian, A. (2010). The effect of visual impairment on quality of life of children aged 3–16 years. British Journal of Ophthalmology, bjo-2010.

[6]. Aziz, N. A. A., Rasli, R. M., and Ramli, K. (2010, December). "Preschool multimedia interactive courseware: Classifying objects (mengelaskan objek) PMICMO". In Software Engineering (WCSE), 2010 Second World Congress on (Vol. 2, pp. 318-322). IEEE.

[7]. Slee, R. (2001). Social justice and the changing directions in educational research: The case of inclusive education. International journal of inclusive education, 5(2-3), 167-177.

[8]. Lyashenko, T.V. (2010) 'Multimedia information technologies in education: basic concepts, essence, and typology (review)', Automatic Documentation and Mathematical Linguistics, Vol. 44, No. 4, pp.206–217.

[9]. Maj, S.P., Tran, B. and Veal, D. (2007) 'State model diagrams – a systems tool for teaching network technologies and network management', Innovations in E-Learning, Instruction Technology, Assessment, and Engineering Education, pp.355–360.

[10]. Quan, Y.S. (2002) 'Network curriculum design and development'[Online].Available:http://etc.Elec.Bnu.Edu.Cn. [Accessed:July. 22, 2015].

[11]. Vodanovich, S., Rohde, M., Dong, C.-s., and Sundaram, D. (2013). Youth Web Spaces: Designing Interfaces as if Youth Mattered. Human IT: Journal for Information Technology Studies as a Human Science, 11(3).

[12]. Zhang, T. (2002) 'Super sports multimedia WCAI web-based teaching courseware development design and application', Sichuan Sports Science, No. 2.

[13]. Shakirat Oluwatosin Haroon and Tosho Abdulrauf. (2015). A Virtual Reality Prototype for Learning Maize Planting. Communications on Applied Electronics 2(1):10-14. Published by Foundation of Computer Science, New York, USA.

[14]. Tosho, A., Abdul Mutalib, A., and Nur Abdul Salam, S. (2014). Usability of Instructional Interface: Accessibility Strategies of Courseware Design for Distance Learning, Nigeria. International Journal of Computer Applications,106(18), 32-36.

[15]. Tosho A., Ariffin A. and Sobihatun N. A. S., (2015). Designing Usability Strategies: Implications for Instructional Interface towards Courseware for Inclusive Education System (C4IES). Research Journal of Applied Sciences, Engineering and Technology. Maxwell Scientific Organization. [Online].Available:http://www.maxwellsci.com/author/ms/attestation/12d1a34af167f99da29f758ec5c11e95-16537-RJASET-DOI.pdf.[Accessed:July. 24,2015].

[16]. Kobayashi, M., Minagawa, H., Nishioka, T., and Miyoshi, S. (2012, July). Meeting support system for the person with hearing impairment using tablet devices and speech recognition. In International Conference on Computers for Handicapped Persons (pp. 221-224). Springer Berlin Heidelberg.

Health Challenges in Using Biometric Technologies and Suggested Solutions

Aro, T. O
Department of Computer Science, University of Ilorin, Ilorin, Nigeria
taiwo774@gmail.com

Oluwade, B. A
Department of Computer Science, University of Ilorin, Ilorin, Nigeria
bamideleoluwade@computer.org

Ogundokun, R. O
Deparment of Computer Science, University of Ilorin, Ilorin,, Nigeria
seunlayan@gmail.com

Oladipo, I. O
Department of Computer Science, University of Ilorin, Ilorin, Nigeria
idoladipo@gmail.com

***ABSTRACT**: A biometric technology is a method of authenticating or detecting of a person based on the uniqueness of individual biological features. This technology has become most effective technique to prevent medical identity thefts and other frauds in healthcare services and insurances, but on other hand its influence is felt adversely on human's health. In this paper, different types of biometric methods are mentioned. The effect of biometrics on human's health vis-a-vis transmission of micro-organisms and germs are discussed, and also the preventive measures to curtail these health hazards are suggested.*

Keywords— *Biological traits, Biometric technology, Health challenges, Health services and Micro-organisms.*

I. INTRODUCTION

As electronic transactions increase, information and digital data become more predominant in healthcare sectors. In the past, people in medical domains have used different techniques to secure their health records on computer systems with methods such as highly encrypted passwords, access codes and personal identification numbers (PINs). All these are conventional methods of authentication, which is either possession based techniques (token like keys, cards or badges) or knowledge based techniques (access to physical or virtual domain based on password and PINs). These methods are subject to several fundamental flaws. They are many a times difficult to remember, easily guessed or forcefully possessed by unauthorized users; the security breaches in these approaches showed that they are unable to satisfy requirements for security of information in health care systems.

The several draw backs in these techniques led to the authentication system based on individual biological characteristics measurement known as biometrics. A biometric is a method that involves the authentication of a person using feature vectors obtained from physiological or behavioural traits [1][2]. Among traits used for recognition task are odour, face, iris, retina, signature, speech, vein and hand geometry. Employing a distinctive feature of the body for identification has been considered to be an suitable and convenient solution in the technological world today.

Although this technology has been in existence for years, but the recent improvements together with drastic reductions in price make these technologies readily accessible and cheaper to users, small enterprise owners, bigger corporations and government. The combination between the biometric data methods and authentication produces the security systems. The security system serves as a lock and it captures the mechanism to control access to confidential data, while the individual biological features are the keys to open the lock [3].

Biometric systems have become the most recent promising technique of recognition in our technological society [4]. Biometric technologies can be used for identification and verification functions [5]. In identification phase, the system only checks its reference dataset to find if a stored reference matches the biological feature vectors stored. In verification, identity is claimed and the matching method is constrained to looking at the reference conforming to this identity. This is illustrated comprehensively in figure 1.1. The biometric technique strengthens the privacy, authentication and minimizes the risks of insecurity in healthcare systems [6]. The technology is becoming a very robust approach for the healthcare provider to resolve fraudulent problems. Biometric system has been adopted recently by several healthcare organizations to secure health records, enable easier access to medical data and guard health care consumers against frauds.

Despite the effective applications of biometric systems in healthcare services, the technologies often create several health challenges in the society. During enrollment phase, the technique of gathering biological features from individual differs depending of what biometric method employed for recognition. Some biometric methods use near infrared, scanner, flash light, ultraviolet beam and contactless devices to capture biometric data. some of these devices serve as reservoirs to microorganisms or germs, which may facilitate the spread of highly infectious diseases from one user to another.

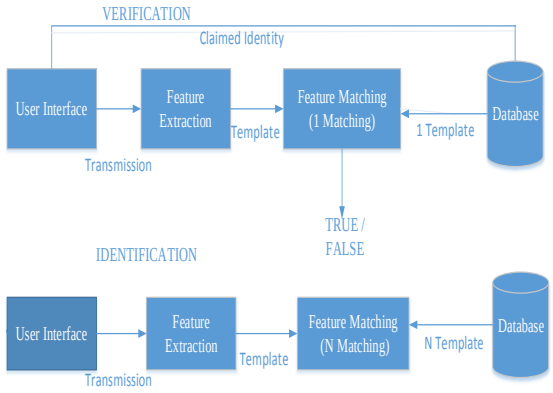


Figure 1.1. Block diagram of two modes of a biometric system [7]

II. BIOMETRIC METHODS OF RECOGNITION

The different techniques require a reliable authentication method to authorize the identity of a person requesting for their services. These biometrical methods include fingerprint, retina scan, iris scan, signature dynamic and body odour, DNA (Deoxyribonucleic acid), speech and face.

(a) Fingerprint

This method is the first commonly used of all biometric methods. It uses unique characteristics of fingerprint to recognize individual. This technology is a widely deployed biometric method in physical and logical access applications [8]. Fingerprint pattern consists of ridges (spaces) and valleys (lines) of human finger. The ridges form the minutia point which is further divided into ridge ends and bifurcation. Fingerprints can be deposited on surfaces such as glasses, metal and polished stone by the human secretions from the eccrine glands that are present in epidermal ridges. The old fingerprint technique uses the ink to obtain a finger point on paper. The paper is scanned using the old scanning device. The new technique uses scanning technologies such as thermal, silicon ,ultrasound and optical with a live finger print scanners [9].

(b) Retina Scan

This technology uses the distinct charateristics of individual blood vessel pattern of the retina for recognition process. Retina is a very small organ that contains neural cells situated in the back region of the eye. The human blood vessels of retina cannot be hereditarily determined, do not change throughout life and thus even they are not the same for twins [1]. Retina cannot be easily seen; it requires an infrared light to obtain the clear structure of this organ. The infrared beam is taken quickly by vessels in the blood of retina than other neighbouring tissues. A retina recognition is performed by sending low intensity infrared illumination into individual's eye as they focus the scanner's eye piece. A coupler is employed to study the features of the blood vessel.. This biometric method is appropriate for organizations that require high security platform, such as government, army and financial institutions. The retina image is showed in figure 2.1.

Figure 2.1 Image of Retina [10].

(c) Iris Scan

Iris of a human being is a small spherical part of the eye, it is accountable for regulating the diameter and dimension of the pupils. It also regulates the amount of light which is allowed to pass through the eye's retina in order to protect it. The iris scan technology uses mathematical pattern verification method on video images of irises of an eye, where complex random patterns are distinctive, stable and can be seen from some distance away. Iris technology uses video camera with near infrared light to acquired images of the detail-rich, complex part of the iris which can be seen outside.

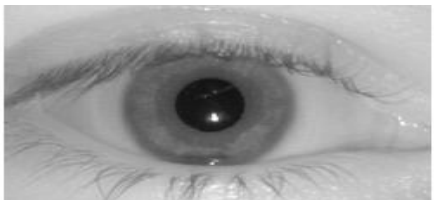


Figure 2.2. Diagram of Human iris [11]

The accuracy of iris-based method is a promising biometric recognition technique. Iris is believed to be unique and even the irises of undistinguishable twins are not having the same form [12].

(d) Vascular (Vein) Recognition

Under the hands are various network of veins, the patterns of these veins are unique for each individual. Vascular recognition employs the scanning technique to obtain vein patterns in the palm, finger or eyeball. This biometric technology works by recognizing the subcutaneous (under the skin) vein strutures in the hand [13]. This process is achieved by placing a user's hand on a scanner applying a very low infrared light to maps the location of the veins. The red blood cells present in the veins absorb the rays and show up on the map as black lines, whereas the remaining hand structure shows up as white. When the vein template is obtained, it is compared with previously stored patterns and a match is performed. There are several parts of the human which can be used for vein recognition which include finger, wrist, palm and backhand. The method of scanning all these human parts are similar.

(e) Signature Dynamic

This technology uses the dynamics of marking that exist in signature rather than comparing the signature itself. The dynamics of signature is computed by using the force, direction, speed, distance of the strokes, number of strokes and their interval. The significant benefit of this technique is that a fraudster cannot gather any information on how to forge the signature by simply studying one that has been written before. Signature dynamic employs several devices to collect the signature dynamics. These devices are traditional tablets and special purpose pens. The tablets gather pressure and 2D coordinates [11]. Special pens are capable of obtaining movements in all three dimensions. In tablet, the final digitalized signature do not look similar to the user's normal signature and also the user does not see what has been written down.

(f) Body odour

Human being emanates special odour that is characterized by its chemical component and this could be used for distinguishing different people. Human body dynamically produces distinctive patterns of volatile organic compounds under various living conditions such as hormonal status, eating, sexual activities, drinking or health [14]. A whiff of air surrounding an individual is blown over an array of chemical sensors, each sensitive to a certain group of aromatic compounds. A component of the odour emitted by a human body is unique to a particular individual. The body odour biometric technology is based on the fact that virtually each human smell is unique. The smell is captured by sensors that are capable of obtaining the odour from non-intrusive parts of the body such as the back of the hand. Each human smell is made up of chemicals known as volatiles. These are extracted by the system and converted into template. The use of the body odour sensors brings up privacy issue as body odour carries an important amount of sensitive personal information. With body odour analysis, diagnosis of some diseases or activities (such as sex) that occurred in past can be deduced [10].

(g) DNA (Deoxyribonucleic Acid)

Deoxyribonucleic acid is a molecule that contains biological information of living organisms, which is also composed of chemical building blocks known as nucleotides. DNA samples that are used for biometric identification can be obtained from several sources like hair, finger nails, mouth swabs, blood , saliva, straws, and any number of other sources that has been attached to human body. The technique is a new biometric technology predominantly used in extreme security domains [16]. It is difficult to fake this method due to the high level of uniqueness (each cell in the human body contains a copy of DNA) and the probability of two persons having the exact same DNA profile except for the fact that is the same in identical twins [15]. It is used generally in the forensic applications for recognition of a particular person.

(h) Speech (Voice)

The voice is a combination of physical and behavioural features that are related to the voice signal patterns of a particular person. Speech recognition in its operation mode normally traces the pattern of voice[5]. There are various speech recognition methods; among those are the acoustics phonetic pattern comparisons and automatic speech recognition method[17].The physical features of voice are related to the appendages that produce its sound. These features include the vocal tracts, mouth, nasal cavities, and lips. Also on the other hand, the behavioural features of voice have connection with the emotional and the speaker physical conditions.

(I) Face Recognition

Face is considered to be unique and distinct part of human that is used to distinguish one person from another. Humans are personally recognized through their faces. Facial recognition is a biometric method for verifying and identifying a person from digital image or video using dataset of face images [18]. The analysis of facial patterns for the authentication or verification of an individual's

identity is achieved either by the use of eigen-faces or local feature analysis. The common techniques for face recognition include the position and form of facial patterns like eyes, nose, lips, chin, and their spatial relationships. Among many biometric methods, the facial biometric system may not be seen as the most dependable and effective technique but its great significant is that it does not require any help from the test subject [19].

III. OPERATION OF A BIOMETRIC TECHNOLOGY

Biometric systems rely on three discrete stages. These include:-

(a) **Enrollment Stage**::- The operation in biometric technologies starts with enrolment. The main duty of enrollment is to gather, store biometric samples and to generate numerical templates for future comparisons. The nature of enrolment phase is determined by a particular method of biometric being employed for identification.

(b) **Template Storage Stage: -** Templates from users are stored in a database for matching by the feature pattern matcher to incoming feature samples. For systems performing only verification (one to one matching), the database may be stored on smart cards, optically read cards or magnetic stripe cards carried by each enrolled user. The database will be centralized for identification process (one to N matching with N greater than one, as in the case of identification or PIN-less verification systems) [11].

(c) **Template Matching Stage: -** The created template is compared to the biometric features stored in the database of system. When a query input template is matched with template in the database, a score is generated. The score number is set based on the level of similarity between the two matching templates. The threshold number is usually configured by a system administrator that creates the level of correlation necessary for a comparison between database template and a query template to be considered for matching. The security level is determined whether the threshold can be set high or low. The best method of setting threshold value depends both upon the numerical characteristics of the comparison spaces coming from the pattern matcher, the relative penalties for false acceptance rate and false rejection rate within the system.

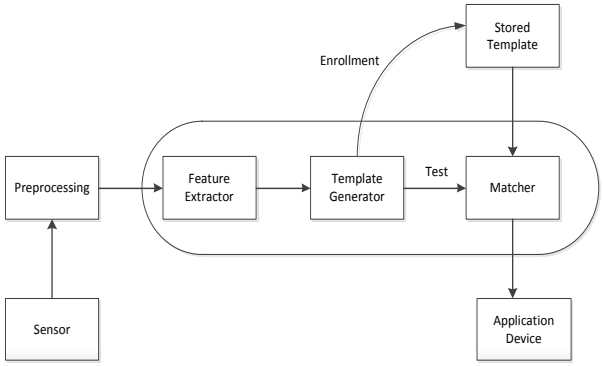


Figure 3.1 A biometric system [15]

IV. HEALTH RELATED CHALLENGES OF BIOMETRIC SECURITY SYSTEMS

Biometric technologies have prompted serious health concerns to many users in the society. The following researchers have worked to review the impact of biometrics on human's health.

[21] presented a study on coming clean on hygiene. He carried out an investigation to show whether fingerprint reader or hand geometry scanner are channels for spread of diseases. The study stated how cultural sensitivities, hygiene requirements and practical considerations right influence the biometric industry's development of verification machines. On the hand, [22] worked on health and safety perceptions of biometric devices. A fingerprint usability study was performed on 300 participants collecting 10 prints. The study reviewed 2% of participants expressed concern about germs and health hazards when touching the biometric scanners. User receptions that are influenced by tactile feedback, duration of contact, usage versus utilization and voluntary versus involuntary were mentioned.

[23] conducted a study to show fingerprint biometric application for visa procurement. They stated that the device and procedure are risk factors for germs transmission. They addressed the risk of microganisms transmission through the use of fingerprint identification method. It was also reviewed in their work that medical personnel have shown that germs possess great potential of transmission through skin of hands. The factors that favour transfer of microorganisms such as large skin surface contact between flat fingers, fingerprint capturing device, non-porous contact surface, high pressure and large overlap of contact surface were analyzed.

The work of [24] focuses on the study of biometrics applications, technology, ethics and health hazards. They

addressed the concept of privacy and data protection, some laws surrounding biometric initiatives, utilization of biodynamic physiological profiles and negative effects from exposing the eye to near infrared light with reference to international legislation.

[25] presented a study on handling the problems in biometrics. Different factors that affect biometric systems such as physical factors, privacy and public confidence, ease of use, and environmental factors were discussed. The physical conditions highlighted a situation in which a user cannot enroll due to some disabilities or limitations in physical characteristics. The study also showed how environmental conditions such as dirty fingers and hands when placed on biometric devices may contribute to the spread of germs from one user to another. Several measures or solutions to prevent those mentioned adverse conditions were recommended in their study.

Hidden risks of biometric identifiers and how to avoid them are the subject of the work of [26]. The work discusses that advances in biometrics will bring about new acceptable contributions and also hazards to information technology world. The presentation mentioned all biometric methods and provides a framework for evaluating them from directions of security, reliability, privacy, potential for abuse and perceived creepiness.

V. SUGGESTED SOLUTIONS TO NEGATIVE EFFECT OF BIOMETRICS TECHNOLOGIES

(a) The biometric devices that involve individual having direct surface contact such as fingerprint, palm print and hand geometry should be properly sanitized or disinfect after each and everyone has used the device in order to minimize the level of micro-organisms. Though, this may consume much time especially at places where there is a high number of people such as at airports, immigration points and other multi-user channels. However, the most paramount thing to be duly considered is our hygiene.

(b) The medical professionals have advised the fingerprint biometric users to wash their hands properly with water and good sanitizer immediately after the use of fingerprint sensors. If unwashed contaminated hands are used, this can facilitate the transmission of bacteria.

(c) Regular cleaning through the use of periodic irradiation with ultra violet light will reduce the level of health concerns and also improve the effective performance of biometrics sensors.

(d) The introduction of electronic smart security system where someone has a personal fingerprint sensor integrated into his/her phone or traveller's document should be considered. At check points the personal fingerprint sensor is used, thus eliminating the unnecessary fear of transmission of pathogens from one person to another.

(e) The option for contactless sensor in fingerprint biometric technology should be envisioned in the nearest future, this will eliminate unnecessary panic of user's health concerns of finger biometric method.

(f) The iris scanning time which involves time interval of about 10 to 15 seconds to be reduced in the future design architecture of iris scanner, without necessarily compromising the accuracy in order to prevent the long period of exposure to near infrared light beam.

(g) During DNA specimen's collection, each user blood or odour samples to be taken with brand new devices. A used device from one user is to be disposed completely just to prevent the transmission of infectious diseases.

VI DISCUSSION

Biometric systems require individuals to supply their physiological or behavioural features to a sensor. The interaction of biometrics users with sensor is critical as many a times how to use a sensor might not be intuitive to inexperienced users. The physiological characteristics as a mean of authentication have created serious health concerns for users. In retinal or iris scan, the user may be asked to keep eyes open for 10 to 15 seconds while being exposed to light beam which may be uncomfortable [25]. The iris scanners do not use lasers but employ digital video camera device with near infra-red light to obtain images of detail rich complex pattern of iris.

The amount of infra-red light used in iris scanning is definitely not more than what would be received from the sun light. The fact is that some people have severe challenges with their eyes which make them to use sun shade when walking in the sun; this same disorder is applicable to light ray from biometric technologies. The regular exposure of eyes to radiations and light beams that are emitted during the scanning may later affect the proper functioning of eyes. Retinal recognition which involves the use of low energy infrared light beam deep inside eye at very close distance and so there is a tendency that the radiation could result into thermal injury on the back of the eye, even though this ocular biometric technique is not currently in use.

Fingerprint is considered currently to be the most dependable biometric characteristic for human verification due to their distinctiveness and persistence[7].The fingerprint biometric is generally used due some advantages attached to this technique such as high accuracy, small memory storage capacity which helps in reducing huge memory dataset required. The technique has been employed in many places such as airports, companies and immigration control system (Border security). The bad news is that biometric devices (sensors or scanners) in which the

fingers are placed can be contaminated by direct contact if they contain germs or pathogens which may unfortunately lead to health problems of the users [8].

In some countries, after the incidence of Ebola, Aids, SARS, Bird-flu and other deadly contagious diseases; it has become a concern to sensitize people how to adequately prevent themselves from any direct contact with other people or a device which involve many users using the same device. The fingerprint sensor for capturing involves the multi-user having direct contact with the same device; this may easily lead to spread of diseases among the users.

Facial recognition seems to offer a great advantage over other methods of biometric systems in term of health perspective. The biometric capturing device in this technology is a digital camera. This method is non-intrusive and contactless (no any direct touch with camera device), since the subject does not need to come very close to the capturing device. The flash light that comes from the camera during face image capturing is neither infrared nor low energy ultra-violet light, this definitely has no direct effect on eyes. Facial recognition is a biometric technique that seems not accommodates any form of health hazardsn [27].

Deoxyribonucleic Acid (DNA) is the highly secured and reliable biometric technique for authentication. This technique may be disastrous to users in an environment where the appropriate medical ethics not duly practised. The blood or odour samples involvement might be collected in an improper way which can result to health concerns of the biometric users.

VII. CONCLUSION

Biometric technologies are very interesting and exciting area that has gained attention in recent years. These technologies have been deployed into several applications in health care for security improvement, effectiveness and reduction of password administrator cost. However, biometric security devices have been identified to be the major and fast channels in which microorganisms, pathogens and diseases can be easily transferred from one user to another during enrollment exercise. This paper exposes different ways by which biometric techniques can be dangerous to humans' health in the society. The preventive measures to minimize these health concerns from biometric users have been suggested in the study.

References

[1] Z. Singhal, P. Gupta, and K. Garg, "Biometric Recognition : Personal Identification Technique,"*International Journal Computational Engineering & Management*, vol. 15, no. 3, pp6–10, 2012.

[2] S. Anila and N. Devarajan, "Preprocessing Technique for Face Recognition," *Global Journal of Computer Science and Technology Graphics & Vision*, vol. 12, no. 11, 2012.

[3] A. K. Jain, A. Ross, S. Pankanti, and S. Member, "Biometrics : A Tool for Information Security," *IEEE Transactions on Iinformation Forensics and Security*, vol. 1, no. 2, pp. 125–143, 2006.

[4] D. N. Parmar and B. B. Mehta, "Face Recognition Methods & Applications," *International Journal Computer Technology & Applications*, vol. 4, no. 1, pp. 84–86, 2013.

[5] A. Srivastva and V. Singh, "Available Online at www.jgrcs.info Biometrics Based Identification Techniques (BIT)," vol. 2, no. 11, pp. 11–15, 2011.

[6] D. G. Mogli, "Role of Biometrics in healthcare privacy and security management system," *Sri Lanka J. Bio-Medical Informatics*, vol. 2, no. 4, pp. 156–165, 2011.

[7] H. Jhaveri, H. Jhaveri, and D. Sanghavi, "Biometric Security System and Its Applications in Healthcare," *International Journal of Technical Research and Applications*, vol. 2, no. 6, pp. 15–20, 2014.

[8] S. Nanavati, M. Thieme, and R. Nanavati, *Biometrics : identity verification in a networked world*. John Wiley & Sons, 2002.

[9] Yi Chen, S. Dass, A. Ross, and A. Jain, "Fingerprint Deformation Models Using Minutiae Locations and Orientations," in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05) - Volume 1*, 2005, pp. 150–155.

[10] D. Bhattacharyya, R. Ranjan, F. A. A, and M. Choi, "Biometric Authentication : A Review," *Int. J. U- e- Source Technol.*, vol. 2, no. 3, pp. 13–28, 2009

[11] J. Wayman, A. Jain, D. Maltoni, and D. Maio, "An Introduction to Biometric Authentication Systems 1," pp. 1–21.

[12] J. Daugman, "How Iris Recognition Works," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 14, no. 4, pp. 458–458, 2004.

[13] B. I. Khan, "Vein Pattern Recognition - Biometrics Underneath the Skin," 2006.

[14] C. Wongchoosuk, M. Lutz, and T. Kerdcharoen, "Detection and Classification of Human Body Odor Using an Electronic Nose," pp. 7234–7249, 2009.

[15] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition" , *IEEE Transaction on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, vol. 14, no. 1, pp. 1–29, 2004.

[16] C. Le, "Abstract : Keyword : Table of Contents : 2- Application Fields for Biometrics Technology," no. Id, pp. 1–10, 2004.

[17] M. N. Uddin, S. Sharmin, A. Hasnat, S. Ahmed, and E. Hasan, "A Survey of Biometrics Security System," *International Journal of Computer Science and Network Security*, vol. 11, no. 10, pp. 16–23, 2011.

[18] A. G. Joshi and A. S Despande, "Review of Face Recognition Techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 1, pp. 71–75, 2015.

[19] S. B. Thorat, H. C. Wolf, and C. Bisson, "Facial Recognition Technology : An analysis with scope in India," vol. 8, no. 1, pp. 325–330, 2010.

[20] M. Kangra and C. Kant, "Biometric System and Its Challenges," *Int. J. Adv. Res.*

Comput. Sci. Softw. Eng., vol. 5, no. 6, pp. 710–716, 2015.

[21] D. Munyan, "Coming clean on hygiene," *Biometric Technol. Today*, vol. 13, no. 4, pp. 7–8, 2005.

[22] M. Theofanos, B. Stanton, S. Orandi, R. Micheals, and N.-F. Zhang, "NISTIR 7403 Usability Testing of Ten-Print Fingerprint Capture," *Nistr*, no. March, pp. 1–56, 2007.

[23] J. A. Jacobs and M. Van Ranst, "Biometric Fingerprinting for Visa Application: Device and Procedure Are Risk Factors for Infection Transmission," vol. 15, no. 5, pp. 335–343, 2008.

[24] C. Nazionale and I. Superiore, "Ethical and social implications of biometric identification technology," pp. 5–11, 2007.

[25] V. Mamillapalli, "Handling the Problems in Biometrics," 2014.

[26] T. P. Keenan, "Hidden Risks of Biometric Identifiers and How to Avoid Them," 2015.

[27] R. Jafri and H. R. Arabnia, "A Survey of Face Recognition Techniques," vol. 5, no. 2, pp. 41–68, 2009.

On the Relationship between Human Immunology and Computer Immunology

Salau-Ibrahim Taofeekat Tosin
Department of Physical Sciences, computer Science Unit
Al-Hikmah University
Ilorin, Nigeria
taofeekatsalau@gmail.com, ttsalau@alhikmah.edu.ng

Jimoh Rasheed Gbenga
Faculty of Communication and Information Science
University of Ilorin
Ilorin, Nigeria
Jimoh_rasheed@yahoo.com
Ilorin, Nigeria
ayisatwuraola@gmail.com

Yusuf-Asaju Ayisat Wuraola
Faculty of Communication and Information Science
University of Ilorin

Abstract*— *In this conceptual paper, the features of computer and human immunology were described with subtle emphasis on computer virus. As would be argued, both are related since the human immunology has been an inspiration* for building more secured, stronger computer systems and interconnection between them over the years. It is expected that the conceptual framework would not only expose researchers to the basics of biological immune system, but also to computer security architectures that can borrow ideas from it. *Review of related research in human and computer immunology was used.

Keywords—Immunology, Human virus, Computer virus

I. INTRODUCTION

In recent times, relationship between computer and human immunology has been the focus of several academic researchers[1]–[3]. This has been in a bid to develop computer information systems that can withstand exposure to increasing dynamic nature of intrusions just like pathogens do in human beings. Human immunology deals with the study of how the body fights against as well as protects itself from bacteria, viruses, protozoa and fungi collectively known as pathogens. This process in the human body is highly coordinated and proven to protect the human body to a large extent despite the constant exposure to these pathogens. Similarly, there are protective mechanisms such as intrusion detection and prevention systems used in computer systems to guard against intrusions. For instance, viruses, worms, Trojans, ping of death, denial of service, penetration attempt and reconnaissance[4].

Globally, every sector of the economy such as government, education, military and financial sector now relies heavily on internet technology to carry out operational and managerial duties. The most alarming part is that threats to the information system assets are increasing geometrically[5]. A recent survey conducted by Cybersecurity Nexus in 2015 revealed that 77% of respondents reported an increase in cyber-related attacks in 2014 compared to the preceding year [6]. In the same survey, 82% opined there would be more attacks in the succeeding year [6]. Also, the report for 2015 revealed that organisations may experience an increased negative impact of cyberattack by 2016[7]. In Nigeria, the use of internet has increased tremendously to 51.1 % as at June 2015 as against 26.5% in 2011[8]. Therefore, seeking inspiration from the defense mechanism of biological organism would be worthwhile since it protects the body effectively against known and new invasions. The use of biological analogies collectively termed computational intelligence is not new to this study. As a result, this present study is one of the earliest study to vividly describe the relationship between human virus and computer virus under the umbrella of human and computer immunology. Detailed review of past studies would be employed. The rest of this paper provides a brief description of human immunology, computer virus and computer immunology. Finally, requirements necessary for designing a computer immune system and possible computer security architectures that can benefit from computer immunology will also be discussed.

II. HUMAN IMMUNOLOGY

The human immune system is a group of cells, molecules, tissues, organs and circulatory systems[9]. The immune system cells are created in the thymus or bone marrow as the case may be. The human body is constantly exposed to pathogens such as virus which is the focus of this study. The body's function is to detect and remove such virus from the body and ensure that new or same virus does not get into the body subsequently [10]. The Human immune system is multilayered in nature, see Fig 1. The layers are made up of physical barriers, physiologic barriers, the innate immune system and adaptive immune system[11], [12]. The physical barrier, for instance, skin and mucous membrane, is the first contact for virus. The physiologic barrier on the other hand, are saliva, sweat and tears. These fluids have varying pH and temperature as well as contain enzymes that make the environment unconducive for thriving. Next is the innate and adaptive immune systems which act on any viruses that evade

the first two layers. Moreover, the adaptive immune system is divided into humoral immunity and cell mediated immunity [13]. The innate and adaptive immunity have cells that make it function effectively, see Fig 2.

A. Innate Immune System

Innate immune system is a form of human immune system that uses a set of receptors called Pattern Recognition Receptors (PRRs)[12]. They recognize molecules specific to pathogens called Pattern

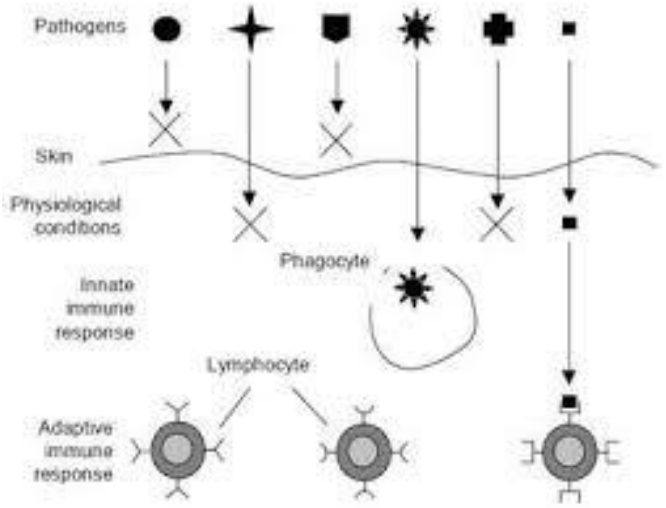


Fig. 1 Layers of Human Immune System [14]

Recognition Receptors (PRRs)[12]. They recognize molecules specific to pathogens called Pathogen Associated Molecular Patterns (PAMPs). This recognition indicates the presence of pathogen in the body. The role of the innate immune system is to defend the host at the onset of infection which is achieved by non-specific recognition of the pathogen. Secondly, it helps to induce the adaptive immune system. Thirdly, the innate immune system determines the type of adaptive response through the expression of co-stimulatory signals in Antigen presenting cells (APCs).

B. Adaptive Immune System

Pathogens that are not destroyed by the innate immune layer get to the adaptive immune layer. This part of the immune system uses complex learning process to achieve better immunity of the human body [11]. Generally, T-cells and B-cells are the lymphocytes used to protect the body. The protection is achieved by using somatically generated antigen receptors clonally distributed on the lymphocytes. This leads to specific recognition of pathogen leading to pathogen-specific long term memory [9]. The receptors used by the adaptive immune system are T-cell receptors (TCR) and B-cell receptors (BCR). Also, the adaptive immune system enables response to pathogen that the body has never encountered before. While B-cell aids production and secretion of antibodies in response to exogenous elements, T-cells attacks infectious cells and control other cell activities in the body [12]. Table 1. give a summary of the major differences between the innate immune system and adaptive immune system.

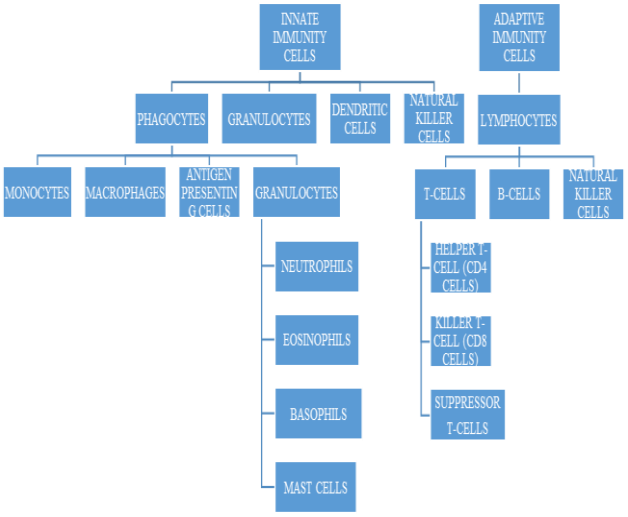


Fig. 2 Cells of the Innate and Adaptive Immune System

III. How Immune System Protects the Body from Pathogens

The pathogen on getting into the body produces antigen (Ag). Antigen is a foreign molecule from pathogen or other exogenous element. Therefore, the target of all processing in the human immune system is directed towards destroying the antigen and maintaining the human body.

The first point of contact by the antigens are phagocytes such as macrophages. Macrophages move freely in the body to digest antigens they encountered. Afterwards, the antigens are fragmented into antigenic peptides [12]. Then, the fragments of the antigenic peptides bind to major histocompatibility complex (MHC) molecules. The antigenic peptide-MHC molecules combination is displayed on the surface of the cell. Next, receptors present on T lymphocyte (T-cell) and B lymphocyte (B-cell) recognize the antigenic peptides – MHC molecule and get activated [13]. T-cells on activation divide and secrete lymphokines (chemical signals) that activate other components of the HIS. Conversely, B-cells on activation respond to the T-cell lymphokines. B-cells have an additional ability of recognizing antigenic peptides that do not bind to MHC molecules. In the same way, B-cells get activated after the recognition process, divide and secrete antibody proteins. These antibody proteins can defuse as well as trigger the destruction of the antigens. Finally, memory cells are developed from the T and B cells which aid elimination of the same antigen if presented in future. As a result, the human immune system is enhanced. More details of how the body is protected can be found in [10].

TABLE I. DIFFERENCES BETWEEN INNATE IMMUNE SYSTEM AND ADAPTIVE IMMUNE SYSTEM

Innate Immune System	Adaptive Immune System

First line of defense	Second line of defense
Evolves and passes from one generation to another	It is an acquired immune system
Response to pathogens is usually within few hours	Response to pathogens takes days or months due to clonal selection of T-cell and B-cell
Referred to as non-specific immunity	Referred to as antigen-specific immunity
The receptors are Pathogen Recognition Receptors (PRR)	The receptors are T-cell (TCR) and B-cell (BCR) receptors

IV. COMPUTER VIRUS

Adleman in 1987 invented the term computer virus [15]. Computer virus, like its biological counterpart human virus, appends parasitically to existing programs in order to attack the resources and create more copies of the virus [16]. When this happens, the host system fails. Subsequently, when virus enters a computer system, it gets executed by the user of the computer system. Afterwards, the virus searches for programs the user has access to and appends itself to the program. Control is immediately returned to the original program, therefore, the user may not notice that another program on the computer system has been infected. As a result, the copy of the virus already in the system gets replicated and spread across the computer system by residing in the memory, in order to monitor system activity. Due to the increasing use of the internet, infected programs are easily transferred to other computer systems electronically.

Most times, computer viruses occur in form of malicious code. The common types of malicious codes are Trojan horse, worms and viruses. Also, the payload contents of these codes generally cause damage [17]. Trojan is a program that disguises as a legitimate program by taking the same name but performs different functions. This similar nomenclature makes it easier for user to mistakenly execute the Trojan. When executed, the intruder circumvents the entry security mechanisms of the computer system. Worms on the other hand execute independently and require a computer network to proliferate. Lapses in network and communication protocol aid their actions. Fig. 3 shows the taxonomy of computer virus. Computer viruses are detected with the use of anti-virus software by employing activity monitors and integrity management systems techniques [15]. The former notify users when system activity associated with viruses occur while the latter warn user when unusual changes are made to files. When a new virus is detected, human experts study and analyze the virus in order to determine the behavior and how it appends itself to the host program. Sequences of instructions present in the virus are collated into signatures that are encoded into the anti-virus software. Analyzing viruses is a difficult and time consuming task that leads to inefficient signatures and consequently high rate of false positives.

According to Harmer [17], virus can also be detected in four different ways which are Appearance, behavior, change and bait. To detect virus by appearance, pattern matching is utilized between scan string and the infected entity. Scan string is the signature made up of sequence of instructions present in the virus. This is used for recognition of the virus in subsequent attacks. As for behavior, the operations performed by the virus is monitored. Moreover, viral activities are similar to normal system operations. Therefore, the antivirus system has to be able to distinguish suspicious activities from normal activities.

When virus gets into the computer system, it infects the host and multiply. This results in changes in the host system. Therefore, it is possible to detect virus by observing the changes in the host. Changes can be easily noticed in the initial state of file. Such changes are file size and file date. Another method is the bait method in which a dummy file usually a replica of file on the computer system is placed in directories or named with file extension prone to virus attack. Such file is also programmed to behave in a virus attractive manner. For instance, executing frequent read and write operations. Afterwards, the original dummy file state is saved and monitored for changes. Occurrence of changes signifies the presence of virus in the computer system. Detailed information regarding the strengths and weaknesses of the detection methods can be found in [17]. The relationship between human virus and computer virus is shown in Table 2 and indicated a compelling reason that supports close analogy between computer virus and human virus.

TABLE II. RELATIONSHIP BETWEEN HUMAN VIRUS AND COMPUTER VIRUS

Human Virus	Computer Virus
DNA contains the definition of self	Program code contains a definition of self
Human virus multiply by reproducing its DNA	Computer Virus multiply by reproducing copy of its program code
Some human viruses attacks specific species	Some computer viruses attacks specific types operating systems
Some attack specific cells	Some attack specific system files. For instance EXE and COM
Not all viral illness in human show symptoms. For instance Hepatitis B virus can be asymptomatic	Also not all computer virus attack shows obvious signs.
Human virus has incubation period before manifestation. Example, Rabies virus has an incubation period up to 6 months	Likewise, some computer virus delays until their payload reaches certain threshold. Example, Michelango virus [17]
Human virus multiplies easily. For instance Human Immunodeficiency Virus (HIV)	Computer virus also makes copies of itself rapidly when it enters the computer system

Consequently, despite the fact that the use of generic anti-virus software removes viruses from computer system, it has few drawbacks. First, it can only detect new viruses that have close sequence to existing viruses. Otherwise, it will be missed. Second, it detects virus only without removing it from the infected file. These two drawbacks necessitated borrowing ideas from the adaptive immune system in human immunology which is specific in its recognition of pathogens. Detection and response to viruses by adaptive immune system is apt such that new viruses can be detected easily. Therefore, designing a computer immune system would be worthwhile in

order to alleviate the problems of generic anti-virus software. The next section is devoted for explaining computer immunology.

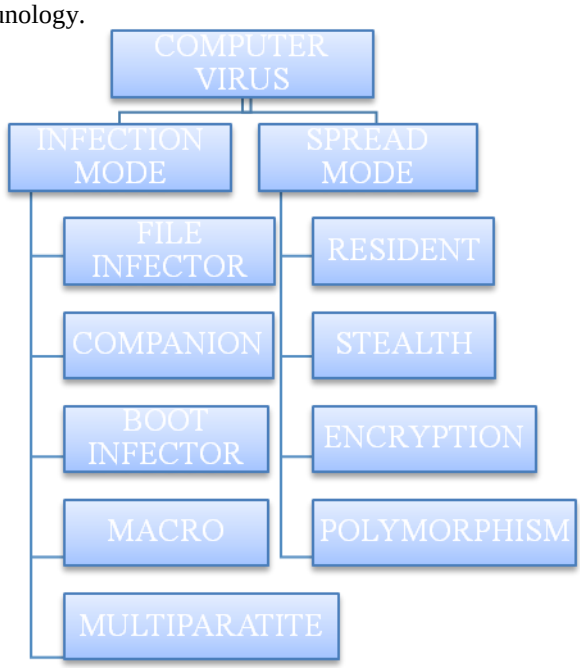


Fig 3. Taxonomy of Computer Virus [17]

V. Computer Immunology

Human immunology and computer immunology are analogous from the security point of view. This is because both ensure protection from invaders. Similar to the Human immune system, computer immune system must incorporate mechanisms from both the innate and adaptive immune system [18]. Human immunology has some desirable features that can be borrowed for improving the security of computer systems. The features according to Kim et al., [19], Ahmad [20] and Somayaji et al., [21] are: distributed, multi-layered, diversity, disposability, autonomy, adaptability, no secure layer, dynamically changing coverage, identity via behavior, anomaly detection, imperfect detection, noise tolerance, self-organizing and generation of antibodies. The listed properties serve as the basis for designing computer immune system for securing computer systems.

Building computer immune system involves certain steps as enumerated by Forrest et al., [2]. First, definition of self which involves defining what constitutes normal activities in a computer system. Detection, prevention and elimination of harmful abnormal activities is important. There should be a memory mechanism for previous intrusion as this aids protection against them when they reoccur. Ways of recognizing new intrusion must be put in place. Finally, protection of the computer immune system from self-attack.

Definition of self in the context of computer system can be normal memory access pattern, network traffic via a router, system call pattern, user behavior pattern, keyboard typing pattern as well as inward and outward flow of TCP/IP packet. Whereas, non-self or intrusions can be in form of misuse, reconnaissance, trojanization, scanning attack, penetration attacks, denial of service attacks and virus attacks. Works such as the one by Burgess explored computer immunology and led to the design of cfengine [1] and also the design of a Computer Virus Immune System (CVIS) by Harmer et al., [4] and Lamont et al., [22]. Extensive review on researches in this area can be found in [19], [23].

A. Computer Immune System Requirements

Kephart et al., [18] outlined the requirements of a computer immune system as innate immunity, adaptive immunity, delivery and dissemination, speed, scalability, safety and reliability, security and finally customer control. Innate immunity involves the ability to detect unknown viruses while adaptive immunity gives the computer immune system ability to automatically detect and remove previously known viruses from the computer system. The process of delivery and dissemination ensure the system successfully remove viruses on the infected host and also from other hosts in a networked environment. Speed allows the removal of viruses within limited period of time before they proliferate. Scalability is concerned with ability to maintain efficiency despite large volume of data. This is in two folds; scalability in virus analysis and updates. As for analysis, computer immune system must be able to handle virus analysis rapidly on one or several computers in a network. In addition, virus analysis must be updated frequently so that re-occurrence of the viruses can be tackled effectively to enhance adaptation. Safety and reliability guarantee low false positive rate as this ensures accuracy in detection. With security, it is important to ensure that scan string produced by the immune system for detecting viruses remain secured. Lastly, customers must have control over the dissemination and delivery of scan string which can be manually or automated.

B. Potential Security architecture for Computer Immune System

This section describes computer security architectures that aid the concepts of computer immunology. Somayaji et al., [21], pointed out four possible architectures. Two additional architectures are added in this paper; heterogeneous and Mobile Ad-hoc Network.

1) Protecting Static Data

This involves protecting data on computer system from unauthorized access by computer viruses. In this architecture, virus- free data is regarded as self while virus- infected data is regarded as non-self. When compared to human immunology, it fits perfectly with the self-non-self-discrimination used for detecting virus and other anomaly in the human body. The mapping between the human components and computer system components is given in the Fig. 4 below.

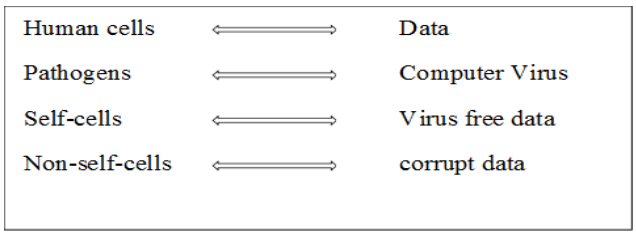


Fig. 4. Mapping Between Human Components and Computer Components.

2) Protecting Active Processes

In this architecture, every active process is regarded as a cell. Also, passwords and file permission represent the skin and the innate immune layer respectively. The idea of lymphocyte can be factor in as "T-cell process". Similar to the T-cell lymphocytes in the human body, T-cell process controls other processes and attack abnormal processes. For that reason, the T-cell process acts by slowing down, suspending, terminating or restarting the attacked process. Also, T-cell process undergoes negative selection in order to generate set of detectors. While some become memory detectors because they detect new intrusion. Moreover, adaptation to changes overtime is achieved by continuous generation of new detectors as well as adjusting their population via aging. See mapping in Fig. 5.

3) Protecting network of Computer

This consist of mutually trusting computers and mutually trusting disposable computers. For the first category, host-based and network-based security mechanism represent the innate immunity while lymphocyte processes i.e. T-cell processes represent the adaptive immunity. The distinguishing features of this architecture are that T-cell processes are mobile agents that serve as detectors. In addition, a computer is dedicated as the thymus. When implemented in a network, it can be used to protect the network and individual computers. In the second category, the main difference from all those discussed so far are the computer components involved. Here, each computer in the network is the cell and the corresponding human component is the network of computers. Network defense mechanism serves

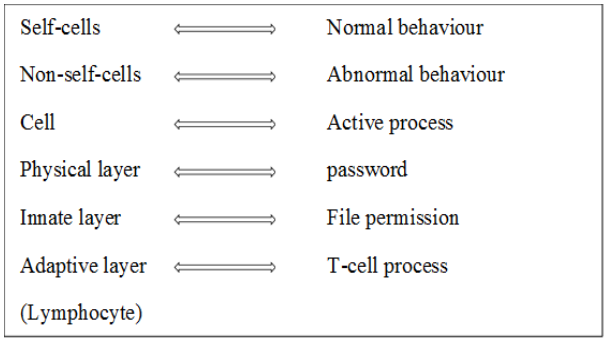


Fig. 5. Mapping Between Human Components and Computer Components

as the innate immunity while a set of lymphocyte machine serves as detectors in the adaptive immunity. The lymphocytes that are negatively selected monitor the computers in the network for the presence of intrusions. Entry of intrusion into the network either results in shutting down or rebooting the affected computer.

4) Protecting Heterogeneous Network

Heterogeneous network consists of computer and other devices with different operating system networked together. This network accommodates interworked architecture among different access network through internet protocol based platforms. This leads to the generation of several types of traffic with varying packet headers, varying MAC protocols and topologies [24]. This means attacks on such network can easily spread across the various access networks. This architecture closely resembles the complex nature of the Human body. Therefore, protecting such network can confidently use the security mechanism employed by the body against external attacks. For instance, individual network can be represented as cell while the whole heterogeneous network is regarded as the human body.

5) Protecting Mobile Ad-hoc Network (MANET)

Mobile Ad-hoc Network is self-organized group of wireless mobile devices (sensor nodes) without fixed infrastructure. Some of its features include but not limited to dynamic network topologies, complex application environment and limited physical security [25]. This type of network is susceptible to different types of security attacks such as viruses attack. Therefore, it is pertinent to provide secure communication for the network. [25]–[27] contains extensive review on Mobile Ad-hoc Network and its security challenges.

VI. Conclusion

The human immune system protects the body from pathogens such as viruses and their dangerous effects. This is achievable because of the various cells and mechanism present. Also, other features of the human immune system provide basis for its adoption in solving computer related problems especially computer virus attack. This paper presents the classification of computer viruses. It describes the requirements necessary to design a computer immune system. Also, the relationship between human and computer immunology was presented. Furthermore, the computer security architectures that can benefit from the human immunology concept were explained.

References

[1] M. Burgess, "Computer Immunology.," 1998, vol. 98, pp. 283–298.

[2] S. Forrest, S. A. Hofmeyr, and A. Somayaji, "Computer immunology," *Commun. ACM*, vol. 40, no. 10, pp. 88–96, 1997.

[3] P. Saurabh, B. Verma, and S. Sharma, "Biologically

Inspired Computer Security System: The Way Ahead," in *Recent Trends in Computer Networks and Distributed Systems Security*, Springer, 2012, pp. 474–484.

[4] P. K. Harmer, P. D. Williams, G. H. Gunsch, and G. B. Lamont, "An artificial immune system architecture for computer security applications," *Evol. Comput. IEEE Trans.*, vol. 6, no. 3, pp. 252–280, 2002.

[5] M. Jouini, L. B. A. Rabai, and A. Ben Aissa, "Classification of security threats in information systems," *Procedia Comput. Sci.*, vol. 32, pp. 489–496, 2014.

[6] ISACA, "State of Cybersecurity: Implication for 2015," ISACA, USA, 2015.

[7] ISACA, "State of Cybersecurity: Implications for 2016," ISACA, USA, 2016.

[8] I. W. Stats, "Nigeria Internet Usage and Telecommunications Reports," *Internet World Stats*. 13-Aug-2016.

[9] J. P. Twycross, *Integrated innate and adaptive artificial immune systems applied to process anomaly detection*. Citeseer, 2007.

[10] C. A. Janeway Jr, "How the immune system recognizes invaders.," *Sci. Am.*, vol. 269, no. 3, p. 72, 1993.

[11] S. Amer, "Enhancing host based intrusion detection systems with danger theory of artificial immune systems," 2008.

[12] L. N. De Castro and F. J. Von Zuben, "Artificial immune systems: Part I–basic theory and applications," *Univ. Estadual Campinas, Dezembro de, Tech. Rep*, vol. 210, 1999.

[13] U. Aickelin, D. Dasgupta, and F. Gu, "Artificial immune systems," in *Search Methodologies*, Springer, 2014, pp. 187–211.

[14] L. N. De Castro and J. Timmis, *Artificial immune systems: a new computational intelligence approach*. Springer Science & Business Media, 2002.

[15] J. O. Kephart, G. B. Sorkin, W. C. Arnold, D. M. Chess, G. J. Tesauro, S. R. White, and T. J. Watson, "Biologically inspired defenses against computer viruses," 1995, pp. 985–996.

[16] T. M. Chen and J.-M. Robert, "The evolution of viruses and worms," *Stat. methods Comput. Secur.*, vol. 1, 2004.

[17] P. K. Harmer, "A distributed agent architecture for a computer virus immune system," DTIC Document, 2000.

[18] J. Kephart, G. Sorkin, M. Swimmer, and S. White, "Blueprint for a computer immune system," in *Artificial immune systems and their applications*, Springer, 1999, pp. 242–261.

[19] J. Kim, P. J. Bentley, U. Aickelin, J. Greensmith, G. Tedesco, and J. Twycross, "Immune system approaches to intrusion detection–a review," *Nat. Comput.*, vol. 6, no. 4, pp. 413–466, 2007.

[20] W. Ahmad, "Natural immune systems for developing novel Machine Learning algorithms," Auckland University of Technology, 2012.

[21] A. Somayaji, S. Hofmeyr, and S. Forrest, "Principles of a computer immune system," 1998, pp. 75–82.

[22] G. B. Lamont, R. E. Marmelstein, and D. A. Van Veldhuizen, "A distributed architecture for a self-adaptive computer virus immune system," 1999, pp. 167–184.

[23] L. N. De Castro and F. J. Von Zuben, "Artificial immune systems: Part II–A survey of applications," *FEEC/Univ. Campinas, Campinas, Brazil*, 2000.

[24] F. Hashim, K. S. Munasinghe, and A. Jamalipour, "On the negative selection and the danger theory inspired security for heterogeneous networks," *IEEE Wirel. Commun.*, vol. 19, no. 3, pp. 74–84, 2012.

[25] L. E. Jim and M. A. Gregory, "A Review of Artificial Immune System Based Security Frameworks for MANET," *Int. J. Commun. Netw. Syst. Sci.*, vol. 9, no. 1, p. 1, 2016.

[26] S. Ghanekar, N. Alrajei, and F. Mili, "Bio-Inspired Intrusion Detection for Wireless Sensor Networks," *Secur. Self-Organizing Networks MANET, WSN, WMN, VANET*, p. 347, 2016.

[27] A.-S. K. Pathan, *Security of self-organizing networks: MANET, WSN, WMN, VANET*. CRC press, 2016.

Mathematical Modeling of the Effect of Electricity Consumption on Economic Growth.

(Case Study: Nigeria)

Adaeze Andra Okeukwu
Enterprise Service and Network Infrastructure
National Identity Management Commission,
Abuja, Nigeria.

Bamidele Oluwade
Department of Computer Science
University of Ilorin,
Ilorin, Nigeria.

***Abstract*— This paper delves into developing a mathematical model to predict the effect of electricity consumption (kWh) on economic growth (in US dollars) in Nigeria for a period of 14 years (2000-2013). The paper adopts the Least Square Regression Method to estimate parameters for the mathematical model. This model makes the assumption that electricity consumed in a country has an effect on her economic growth and represents economic growth as Gross Domestic Product (GDP). The mathematical model is developed as a prediction tool for future economic growth of Nigeria. This work reveals that Nigeria needs to increase electricity produced in order to meet demand in the country. Moreover, from the mathematical model, Nigeria needs about 500 kWh per capita of electricity consumption to increase its GDP to 18828.03 US dollars per capita.**

Keywords— economic growth; electricity consumption; mathematical model; Nigeria

I. INTRODUCTION

In the last 100 years, there has been an intense increase in the advancement and use of technology. These development had led to increased energy demands in most countries. The developed nations, however, using different approaches, have met their nation's energy demands and are presently, focusing on delivering cleaner and more efficient energy alternatives.

However, Nigeria has been slow in following this trend. Over the years, numerous propositions and efforts has been put forward to address this energy crisis. But the problem seem to mystify all concerted efforts. In 2011 and 2012 alone, the Federal Government of Nigeria spent 8.57% and 5.33% respectively of its Federal budget on Power alone which was a total of 156.55 billion naira. Yet, improvements have not been visible. Fatih Birol, IEA Chief Economist said "Economic and social development in sub-Saharan Africa hinges critically on fixing the energy sector. The payoff can be huge; with each additional dollar invested in the power sector boosting the overall economy by $15."

Whether in developed or developing countries, the power sector is the key sector in every economy. Steady and adequate supply of power (electricity) is hallmark for economic growth and development, especially in the real sector. As Alam (2006), puts it, "energy is indispensable force driving all economic activities". In other words, an increase in energy (electricity) consumption will push economic growth.

Furthermore, many research has been conducted to observe if there exist a causal relationship between electricity consumed and economic growth (Gross Domestic Product). Some works had gone further to observe the direction of the causal relationship between electricity consumption and economic activity. If a unidirectional causality runs from economic growth to electricity consumption it implies that policies for increasing electricity consumption may be implemented with an increasing effect on economic growth but not vice versa. In contrast, if bidirectional causality is found, economic growth may demand more electricity whereas more electricity consumption may induce economic growth. Both electricity consumption and economic growth complement each other and energy conservation measures may negatively affect economic growth. Yoo (2005) for Korea and Yoo (2006) for Malaysia found bidirectional causality between electricity consumption and economic growth. Lastly, if no causality in any direction, this indicates policies for increasing or reducing electricity consumption do not affect economic growth, and rise in economic growth may not affect electricity consumption.

Many economists do agree that a strong correlation between electricity and economic growth exist. Morimoto and Hope (2001) has discovered, using Pearson correlation coefficient that economic growth and electricity consumption in Sri Lanka are highly correlated. Akinlo (2009) investigated the causality relationship between energy consumption and economic growth for Nigeria from 1980–2006, he reported that real gross domestic product and electricity consumption are cointegrated and there is only unidirectional Granger causality running from electricity consumption.

Also, Omisakin (2008) investigated the relationship that exists between energy consumption and economic growth using Nigerian annual time series from 1970-2005. His findings revealed a unidirectional causal relationship between total energy consumption and economic growth running from total energy consumption to economic growth. Therefore, it is

on this basis that a model has been developed for predicting the economic growth using electricity consumption.

II. ELECTRICITY CONSUMPTION PATTERN IN NIGERIA

For any economy to prosper, producers will have to supply what consumers demand. Yet, Nigeria has been unable to meet the energy demands of its increasing population. In 2012, based on World Bank Group Statistics, the population of Nigeria was estimated to be 168.8 million, yet only 55.6 % of that population had access to electrical energy. This shows that a massive number of the population have little or no access to electrical energy leading individuals and organizations to meet their electricity needs using fuel generators and solar power cells.

The trend of Electricity Consumption (kWh per capita) in Nigeria from 2010 to 2013 is shown in Fig. 1. It can be assumed that demand for electricity is equivalent to the population of Nigeria which can be observed keeps increasing. Also electricity consumption is assumed to be the supply. The pattern shows an inconsistent electricity consumption over this period. From the graph, we can observe a constant increase in the Nigerian population and an erratic pattern for electricity consumption. In 2001,2003,2006,2009 and 2013, there was sharp decline in electricity consumption with demand increasing. It is evident, observing this two trends that supply does not meet demand.

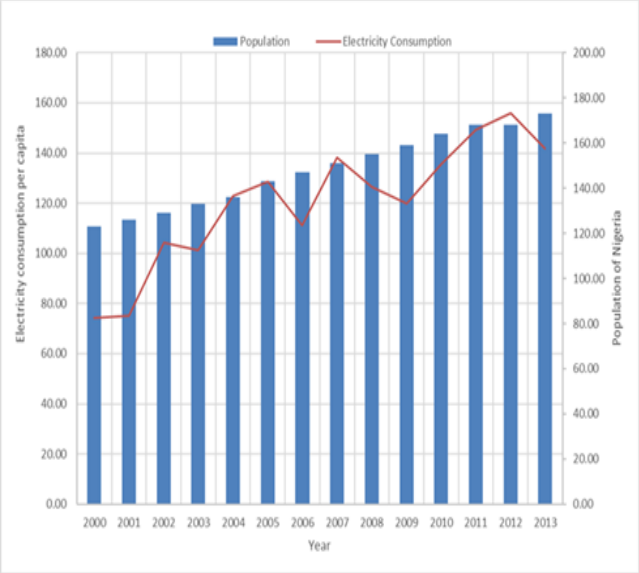


Fig 1. Pattern of Electricity Consumption and Population in Nigeria from 2000 to 2013.

The Gross Domestic Product (GDP) of a nation indicates the economic growth and health of that nation. It indicates the standard of living [of people living] in a particular country. In modelling GDP against electricity consumption, assumptions have been made that the economic growth of a nation is affected by electricity consumed. In Fig. 2, we observe the trend of GDP and electricity growth from 2000 to 2013. It can be observed that Nigeria has a more stable GDP compared to electricity consumed. This can be related to other factors such as labour force affecting GDP and Nigeria has massive labour force. Hence, labour has been included in this research.

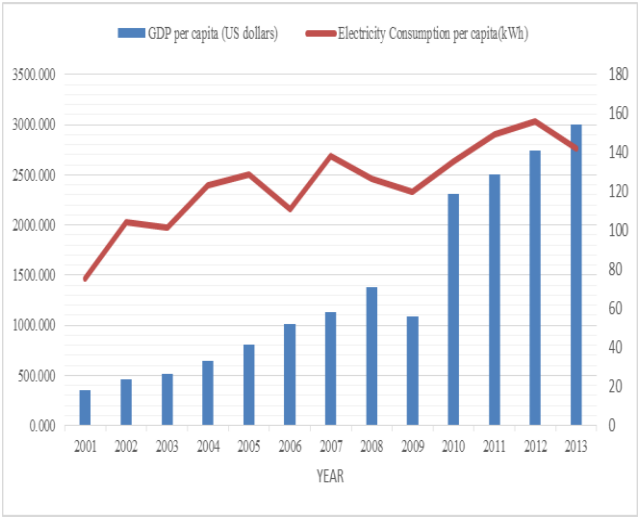


Fig 2. Pattern of GDP and Electricity Consumption in Nigeria from 2000 to 2013.

III. MATHEMATICAL MODEL

In formulating this empirical model, we assume that:

- Nigeria has put down a good policy to control price of electricity and electricity consumed is negligent of the price of electricity.
- There exist a causal relationship between economic activity and electricity consumed.
- Nigeria uses the low cost of labor to substitute and reduce the effect of its inadequate power supply on its economy.

TABLE 1. DATA FOR GDP, ELECTRICITY CONSUMPTION AND LABOR FORCE

Year	GDP per capita (US Dollars)	Electricity Consumed (kWh per capita)	Labor Force
2000	377.500	74.131	38.876×10^6
2001	350.290	75.197	39.626×10^6
2002	457.470	104.135	40.482×10^6
2003	510.420	101.402	41.222×10^6
2004	645.930	122.985	42.064×10^6
2005	804.150	128.636	43.250×10^6
2006	1014.760	111.144	44.460×10^6
2007	1130.880	138.142	45.660×10^6
2008	1376.020	126.532	47.008×10^6
2009	1090.750	119.949	48.330×10^6

Year	GDP per capita (US Dollars)	Electricity Consumed (kWh per capita)	Labor Force
2010	2310.860	135.638	49.707×10^6
2011	2507.680	149.312	51.167×10^6
2012	2742.220	155.854	52.601×10^6
2013	3005.510	141.873	54.199×10^6

[a.] Source: Compiled by author from World Development Indicators of World Bank, 2016.

Let the equation be of the form:

$$G = a + bE + cE^2 + dL + eL^2 + fEL \quad (1)$$

Where G = Economic growth, E = Electricity Consumption, L = Labor Force

Using the Least Squares Method,

$$I = \min \sum_{i=1}^{14} (G_i - a - bE_i - cEi^2 - dLj - eLj^2 - fE_iL_j)^2 \quad (2)$$

$$\partial I / \partial a = 2 \sum_{i=1}^{14} - (G_i - a - bE_i - cE_i^2 - dL_j - eL_j^2 - fE_iL_j) = 0$$

$$\sum_{i=1}^{14} G_i = a\sum_{i=1}^{14} + b\sum_{i=1}^{14} E_i + c\sum_{i=1}^{14} E_i^2 + d\sum_{j=1}^{14} L_j + e\sum_{j=1}^{14} L_j^2 + f\sum_{j=1}^{14} E_iL_j \quad (3)$$

$$\partial I / \partial b = 2 \sum_{i=1}^{14} (-E_i)(G_i - a - bE_i - cE_i^2 - dL_j - eL_j^2 - fE_iL_j) = 0$$

$$\sum_{i=1}^{14} G_iE_i = a\sum_{i=1}^{14} E_i + b\sum_{i=1}^{14} E_i^2 + c\sum_{i=1}^{14} E_i^3 + d\sum_{\substack{i=1\\j=1}}^{14} E_iL_j + e\sum_{\substack{i=1\\j=1}}^{14} E_iL_j^2 + f\sum_{\substack{i=1\\j=1}}^{14} E_i^2L_j \quad (4)$$

$$\partial I / \partial c = 2 \sum_{i=1}^{14} (-E_i^2)(G_i - a - bE_i - cE_i^2 - dL - eL^2 - fE_iL) = 0$$

$$\sum_{i=1}^{14} G_iE_i^2 = a\sum_{i=1}^{14} E_i^2 + b\sum_{i=1}^{14} E_i^3 + c\sum_{i=1}^{14} E_i^4 + d\sum_{\substack{i=1\\j=1}}^{14} E_i^2L_j + e\sum_{\substack{i=1\\j=1}}^{14} E_i^2L_j^2 + f\sum_{\substack{i=1\\j=1}}^{14} E_i^3L_j \quad (5)$$

$$\partial I / \partial d = 2 \sum_{\substack{i=1\\j=1}}^{14} (-L_j)(G_i - a - bE_i - cE_i^2 - dL_j - eL_j^2 - fE_iL_j) = 0$$

$$\sum_{\substack{i=1\\j=1}}^{14} G_iL_j = a\sum_{\substack{i=1\\j=1}}^{14} L_j + b\sum_{\substack{i=1\\j=1}}^{14} E_iL_j + c\sum_{\substack{i=1\\j=1}}^{14} E_i^2L_j + d\sum_{\substack{i=1\\j=1}}^{14} L_j^2 + e\sum_{\substack{i=1\\j=1}}^{14} L_j^3 + f\sum_{\substack{i=1\\j=1}}^{14} E_iL_j^2 \quad (6)$$

$$\partial I / \partial e = 2 \sum_{\substack{i=1\\j=1}}^{14} (-L_j^2)(G_i - a - bE_i - cE_i^2 - dL_j - eL_j^2 - fE_iL_j) = 0$$

$$\sum_{\substack{i=1\\j=1}}^{14} G_iL_j^2 = a\sum_{\substack{i=1\\j=1}}^{14} L_j^2 + b\sum_{\substack{i=1\\j=1}}^{14} E_iL_j^2 + c\sum_{\substack{i=1\\j=1}}^{14} E_i^2L_j^2 + d\sum_{\substack{i=1\\j=1}}^{14} L_j^3 + e\sum_{\substack{i=1\\j=1}}^{14} L_j^4 + f\sum_{\substack{i=1\\j=1}}^{14} E_iL_j^3 \quad (7)$$

$$\partial I / \partial f = 2 \sum_{\substack{i=1\\j=1}}^{14} (-E_iL_j)(G_i - a - bE_i - cE_i^2 - dL_j - eL_j^2 - fE_iL_j) = 0$$

$$\sum_{\substack{i=1\\j=1}}^{14} G_iE_iL_j = a\sum_{\substack{i=1\\j=1}}^{14} E_iL_j + b\sum_{\substack{i=1\\j=1}}^{14} E_i^2L_j + c\sum_{\substack{i=1\\j=1}}^{14} E_i^3L_j + d\sum_{\substack{i=1\\j=1}}^{14} E_iL_j^2 + e\sum_{\substack{i=1\\j=1}}^{14} E_iL_j^3 + f\sum_{\substack{i=1\\j=1}}^{14} E_i^2L_j^2 \quad (8)$$

Substituting values from Table 2 into Equations (3), (4), (5), (6), (7) and (8) above,

$18324 = a + 1684.93b + 210893.457c + (638.652\times10^{6})\ d + (29459.834\times10^{12})\ e + (78257.595\times10^{6})\ f$ (9)

$2452209.063 = 1684.93a + 210893.457b + 27219592.408c + (78257.595\times10^{6})d + (3672805.079\times10^{12})e + (9944816.489\times10^{6})\ f$ (10)

$334706319.300 = 210893.457a + 27219592.408b + 3598846172.861c + (9944816.489\times10^{6})\ d + (473490183.095\times10^{12})\ e + (1300005161.257\times10^{6})\ f$ (11)

$894532.595\times10^{6} = (638.652\times10^{6})\ a + (78257.595\times10^{6})\ b + (9944816.489\times10^{6})\ c + (2945.834\times10^{12})\ d + (1374058.260\times10^{18})\ e + (3672805.079\times10^{12})\ f$ (12)

$44027623.994\times10^{12} = (29459.834\times10^{12})\ a + (3672805.079\times10^{12})\ b + (473490183.095\times10^{12})\ c + (1374058.260\times10^{18})\ d + (64788824.622\times10^{24})\ e + (174139002.734\times10^{18})\ f$ (13)

$120994255.840\times10^{6} = (78257.595\times10^{6})\ a + (9944816.489\times10^{6})\ b + (1300005161.257\times10^{6})\ c + (3672805.174\times10^{12})\ d + (174139002.735\times10^{18})\ e + (473490183.095\times10^{12})\ f$ (14)

Solution:

$a = -1.565508653742$

$b = -47.0305480163867$

$c = 0.0832848239156951$

$d = -1.08788562978169\times10^{-8}$

$e = 8.54006816276737\times10^{-13}$

$f = 7.00923291951894\times10^{-7}$

Substitute the values of a, b, c, d, e, f into Equation 9 above;

$a + 1684.93b + 210893.457c + (638.652\times10^{6})\ d + (29459.834\times10^{12})\ e + (78257.595\times10^{6})\ f = G$

Therefore G = 18324.0

Also, substitute the values of a, b, c, d, e, f into Equation 1 above; the equation is,

$G = -1.56551 - (47.0305)\ E + (0.0832848)\ E^{2} - (1.0879\times10^{-8})\ L + (8.54007\times10^{-13})\ L^{2} + (7.0092\times10^{-7})\ EL + e_i$ (15)

Where e_i = Residual term

$e_i = \sum (Y - Y')$ (16)

Where Y= Real value of GDP per capita, Y'= Predicted value of GDP per capita.

From Table 3 below, e_i =20.92

Finally, Mathematical Model is

$G = -1.56551 - (47.0305)\ E + (0.0832848)\ E^{2} - (1.0879\times10^{-8})\ L + (8.54007\times10^{-13})\ L^{2} + (7.0092\times10^{-7})\ EL + 20.92$

TABLE 2. SUMMATION OF VARIABLES USED IN MODELING EQUATION

No	Variable	Total(2000-2013)
1	GDP per capita (G)	18324.440
2	Electricity Consumption per Capita (E)	1684.930
3	Labor Force (L)	638.652
4	E^2	210893.457
5	E^3	27219592.408
6	E^4	3598846172.861
7	GE	2452209.063
8	GE^2	334706319.300
9	L^2	29459.833×10^{12}
10	L^3	1374058.259×10^{18}
11	L^4	$64788824.622\times10^{24}$
12	GL	894532.596×10^{6}
13	GL^2	$44027623.992\times10^{12}$
14	GLE	$120994255.840\times10^{6}$
15	EL	78257.596×10^{6}
16	EL^2	3672805.079×10^{12}
17	E^2L	9944816.489×10^{6}
18	E^2L^2	$473490183.095\times10^{12}$
19	E^3L	$1300006333.126\times10^{6}$
20	EL^3	$174139002.735\times10^{18}$

TABLE 3. GDP PER CAPITA (REAL AND PREDICTED VALUES)

Year	Real GDP per capita (Y)	Predicted GDP per capita (Y)	e_i = Y-Y'
2000	377.500	279.968	97.532
2001	350.290	361.942	-11.652
2002	457.470	357.955	99.515
2003	510.420	466.379	44.041
2004	645.930	610.73	35.2

Year	Real GDP per capita (Y)	Predicted GDP per capita (Y)	e_i = Y-Y'
2005	804.150	823.327	-19.177
2006	1014.760	951.287	63.473
2007	1130.880	1291.95	-161.07
2008	1376.020	1436.71	-60.69
2009	1090.750	1613.04	-522.29
2010	2310.860	1986.8	324.06
2011	2507.680	2423.18	84.5
2012	2742.220	2800.12	-57.9

From mathematical predictions made in Table 4, Nigeria's GDP can be expected to rise if there is an increase in existing power supply. If electricity consumption can be increased to 500 kWh per capita. This can increase GDP to over 18,000 US dollars per capita.

TABLE 4. FUTURE PREDICTIONS BASED ON MATHEMATICAL MODEL

GDP per capita(US Dollars)	Electricity consumption(kWh per capita)	Labor Force(constant)
3045.16	150	54.199×10^6
4050.59	200	54.199×10^6
5472.43	250	54.199×10^6
7310.71	300	54.199×10^6
12236.53	400	54.199×10^6
18828.03	500	54.199×10^6

IV. IMPLICATION FOR ECONOMIC GROWTH

The United States with a population of about 316.5 million in 2013 had the value of electricity consumption per capita to be 12,988.256 kWh. From Table 4 below, a comparison is made for electricity consumption per capita, GDP per capita and population of 8 different countries. Nigeria is trailing behind the industrialized world. Looking at Table 5, it is observed that Japan, USA, Germany and UK all generate sufficient electricity for their citizens and high GDP per capita. It shows that the availability of electricity can stimulate economic growth.

Table 5. Comparison of GDP per capita between countries.

Country	Population	GDP per capita(US Dollars)	Electricity Consumption(kWh per capita)
Nigeria	173.6 million	3,005.51	142
USA	316.5 million	53,041.98	12,988
Germany	80.62 million	46,268.64	7,019
United Kingdom	64.1 million	41,787.47	5,407
South Africa	52.98 million	6,617.91	4,326
Japan	127.3 million	38,633.71	7,836
Russia	143.5 million	14,611.70	6,539

The achievement of a better standard of living for Nigeria requires a total overhauling of the Power sector. Every citizen in Nigeria has a right to electricity in order to be productive. The Nigeria industries need to be powered in order to run effectively and efficiently.

Nigeria has a target to become industrialized and can do that if the electricity consumption is increased. To achieve this objective this, requires a quantum leap and sincere efforts from all those involved.

V. CONCLUSION

The conclusion of this study is that there exist a correlation between electricity consumption and economic growth. This is obvious for Nigeria. But there are anomalies; situations where electricity supply rises and GDP per capita dips. That can be due to a higher increase in population or other factors affecting GDP per capita.

Hence, it can be concluded that electricity supply has an effect on economic growth until demand is fully met. This is because it is electricity demand, not supply that counts. The "effect" as already stated means it has an influence and not that it is the only factor. Hence, the anomalies can be attributed to other factors.

Therefore, massive improvement in electricity supplied and if adequately utilized will induce economic growth in Nigeria. So far, Nigeria has invested a lot into the power sector (finance, good policy, recruitment of experts, etc.) and this has failed to improve the state of electricity consumed in the country. The country needs to discover the reasons for such policy failure and tackle them holistically.

Massive corruption has been highlighted as one of the major factors that impedes growth in the power sector. Hence, government and citizens should exert efforts to curb corruption and punish those who are found guilty of corruption. Also, Nigeria should look beyond fossil fuel as a means of generating electricity and also focus on renewable

means such as wind, solar, hydro in order to increase electricity output for its citizens.

Acknowledgment

The authors want to thank Oguntade Philip Temitope for his numerous contributions to this work.

References

[1] Akinlo, A.E. (2009). "Electricity consumption and economic growth in Nigeria: Evidence from cointegration and co-feature analysis." Journal of Policy Modeling, Volume 31, Issue 5, Pages 681–693

[2] Alam, M. S. (2006). "Economic growth with energy." Munich Personal RePEc Archive (MPRA), (1260), 1–25.

[3] Iwayemi, A. (2007)."Investment in Electricity Generation and Transmission in Nigeria: Issues and Options." International Association for Energy Economics.

[4] Enebeli E. E. (2010)." Causality Analysis of Nigerian Electricity Consumption and Economic Growth" Journal of Economics and Engineering, ISSN: 2078-0346, №4.

[5] Hussain, A. B. and Salwati bt Othman, N. (2011). "Causality analysis among electricity consumption, consumer expenditure, gross domestic product (GDP) and foreign direct investment (FDI): Case study of Malaysia." Journal of Economics and International Finance Vol. 3(4), pp. 228-235.

[6] Babatunde, M. A. and Shuaibu, M. I. (2007). "The Demand for Residential Electricity in Nigeria: A Bound Testing Approach."

[7] Adeboye, K. R. (2010). "Mathematical Modeling" Mat 416 Lecture Note (unpublished). Mathematics & Computer Science Department, Federal University of Technology Minna, Nigeria.

[8] Ayodele, S. (2001). "Improving and Sustaining Power (Electricity) Supply for Socio-Economic Development in Nigeria." International Association for Energy Economics.

[9] Sheng-Tung, C., Hsiao-I, K. and Chi-Chung, C. (2007)."The relationship between GDP and electricity consumption in 10 Asian countries." Energy Policy 35, pp 2611–2621.

[10] Wikipedia (2011)."Gross Domestic Product, Electricity".Retrieved from https://en.wikipedia.org/wiki/.

[11] Yoo, S. H. (2005). "Electricity consumption and economic growth: Evidence from Korea." Energy Policy, 33, pp.1627-1632.

[12] Yoo, S. (2006). "The causal relationship between electricity consumption and economic growth in ASEAN countries." Energy Policy, 34, pp.3573-3582.

[13] Morimoto, R. and Hope, C. (2001). "The impact of Electricity Supply on Economic Growth in Sri Lanka." The Judge Institute of Management Studies, Trumpington Street, Cambridge CB21AG.

[14] Omisakin, A. (2008). "Energy Consumption and Economic Growth in Nigeria: A Bounds Testing Cointegration Approach". Journal of Economic Theory 2 (4): 118-123, 2008.

[15] Federal Republic of Nigeria (1999-2012). "Government Spending Summary". www.fmf.gov.ng.

[16] Birol, F. (2014, October 13). "Energy sector is key to powering prosperity in sub-Saharan Africa. International Energy Agency." Retrievedfromhttps://www.iea.org/newsroomandevents/pressreleases/2014/october/energy-sector-is-key-to-powering-prosperity-in-sub-saharan-africa.html.

[17] The World Bank, World Development Indicators (2016). GDP per capita, Electricity Consumption per capita, Labor force. Retrieved from http://data.worldbank.org/indicator/.

Application of Metaheuristic Algorithms to Personnel Scheduling Problem in Healthcare Management: A Literature Review

Agboola, O. M.
Department of Computer Science
Kwara State Polytechnic
Ilorin, Nigeria
oladiranagb@gmail.coom

Sadiku, J. S. and Bajeh, A. O.
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
jssadiku@unilorin.edu.ng, bajehamos@unilorin.edu.ng, bajeh_amos@yahoo.com

***Abstract*— This paper presents a literature review to evaluate the use of metaheuristic algorithm in health care staff and space scheduling to ensure maximum productivity in controlling and management of diseases. The paper reviewed several studies in the application of algorithms to enhance the productivity in the area of healthcare through optimal scheduling or proper nurse, physician, patient and space. The result of the review shows that proper application of ICT in health care service will enhance the service delivery in terms of optimal scheduling of medical staff and space, controlling and management of diseases.**

Keywords*— Metaheuristic Algorithm, Heuristic Algorithm, Scheduling, Healthcare

I. INTRODUCTION

Personnel scheduling problem is faced by multi-staffed and space organizations some especially in healthcare industry. Creating a schedule that adequately covers patient's demand and satisfying the preferences of employees as well as observing work regulations and ensuring fair distribution of work is a complex [1].

As a scheduler must endeavor to satisfy staff and patients with competing preferences to work schedules, they must also consider the problem of balancing the workload throughout the planning period. These tasks must be carried out properly so that adequate service would be given to their targeted audience (patient) in controlling and management of diseases. In view of this, there is need for adequate and optimal scheduling of medical staff in order to enhance medical service delivery. (Figure 1 depicts the pseudo code for the physician greedy assignment algorithm prepared in [1]). This necessitated the review of algorithm that can be used to enhance service delivery in healthcare establishments.

Space Allocation Problem (SAP) is the process of planning the utilization of space by determining the position and size of several elements according to some specified requirements. The allocation of spaces to entities is systematically done to optimize the use of the available space [2]. SAP seeks to allocate given entities/goods into available scarce resources while satisfying given requirements and constraints. It is a multi-objective combinatorial optimization problem similar to bin-packing and Knapsack problem [3].

Health care personnel scheduling problem is a combinatorial optimization problem, solving many versions of allocation problem which may consume heavy amount of processing power.

However, they can be efficiently implemented in terms of memory requirements and time. The time complexity concept includes decision and optimization problem which includes the classes of P, NP, NP-complete, NP-hard and no free lunch theorem.

The Medical Space Allocation also belongs to the combinatorial optimization problem. The medical schedule allocation problem is said to be finite in quantity and must be allocated and properly managed to ensure maximum productivity. For instance entities with similar interest/need to access specialized equipment will be grouped together. There are two type of space allocation involve in health care centers. First is the clinical space which includes rooms such as waiting room, consulting room, examining room, labs, medical/utility support room, medical staff room and special care room required for primary health care staff to discharge their primary assignment. The second type includes facility support space such as the reception, work areas, counseling and meeting room, kitchens, ceremonial space and storage [4].

The objectives of this paper are to review the existing studies in which heuristic/metaheuristic algorithms are used in the scheduling and allocation of medical personnel space in healthcare delivery. It also highlights the merit of ICT algorithms on scheduling allocation problem.

II. METAHEURISTIC ALGORITHMS

Metaheuristic algorithms are improved heuristic algorithm that build upon on heuristic algorithm techniques to make them to be more effective and efficient in searching the solution space, preventing premature convergence and local minimum. This section presents some of the most commonly used metaheuristic algorithms.

A. Simulated Annealing (SA)

This is a genetic probability metaheuristic algorithm for global optimization problem of locating good approximation to the global optimum of a given function in a large space. It is often used when the search space is discrete.

B. Tabu Search Algorithm

Tabu Search is another metaheuristic algorithm that attempts to search a given space in a systematically and intelligently, using flexible and adaptive memory structure and intensification and exploration strategies. Its main components are long and short term memory and diversification strategies.

C. Variable Neighbourhood Search (VNS)

This algorithm is based on the idea of using multiple neighbourhood structures during the local search. UNS algorithm tries to exploit the motion that a local optimum with respect to a single neighbourhood structure may not necessarily be optimal from the point at another neighbourhood structure

D. Greedy Randomize Adaptive Search Procedure Algorithm (GRASPA)

GRASPA is an iterative multi-start metaheuristic in which each iteration of the space search consists of two stages: construction and local search. A feasible solution is generated in the construction stage and this solution is improved to a local optimal by means of a local search operator.

E. Genetic Algorithm (GA)

GA is basically the mimicking of the mechanism of the natural biological genes selection by artificial system. The algorithm involves the generation of a population of individual which iteratively evolves into a best fit population through the means of self-adaptation and recombination. Brememan, (1962) introduced the idea of using evolution and recombination for optimization in GA.

F. Particle Swarm Optimization (PSO)

Like the GA, PSO is also a biological inspired computation search and optimization method developed in 1995 by Eberhart and Kennedy based on the social behavior of birds flocking or first schooling.

G. Artificial Bee Colony (ABC)

Artificial Bee Colony (ABC) Algorithm is a swarm based metaheuristic algorithm modeled after the foraging behavior of honey bee colonies and used for the optimization of numerical problems. This algorithm was introduced by Karaboga in 2005 inspired by the intelligent foraging behavior of honeybees.

III. LITERATURE REVIEW

This section presents the review of the studies in the use of metaheuritic for allocation of medical personnel and space in healthcare delivery organizations.

In [5], Mohammed, et al. proposed solving health care facility location problem using an Improved Particle Swarm Optimization algorithm (IPSO) to maximize the population assigned to facility within a given coverage distance. The proposed IPSO approach was compared with Lingo 8 software and Genetic Algorithm (GA) in facility location. The results of the comparison showed that IPSO produces better facility location at a shorter time than each of Lingo 8 and GA.

In [6], Group Genetic Algorithm (GGA) was proposed by Muting and Mbohwa for assigning care tasks to healthcare givers at a minimal cost while considering healthcare service quality by striving to meet the time window restriction appropriate for patients. The approach utilizes the strengths of unique group genetic operators to effectively and efficiently address the group structure of the problem. This algorithm provides a good solution within reasonable computation time when compared to PSO, GA and Parallel PSO.

Gerhard, et al. in [7], provided a generic framework for solving a real-world multi-modal home-healthcare scheduling using variable neighborhood search, memetic algorithm, scatter search and simulated annealing to assign home-care staff to customer and determining efficient multimodal tours, while considering staff and customer satisfaction. A comparative study showed that the variable neighborhood search approach is capable of solving world instances in reasonable time and produces valid solution within a few seconds.

Masri, et al. [8] These researchers used an Enhanced Basic Harmony Search Algorithm (EHSA) using a semi cyclic shift patterns in the initialization step to generate the initial harmonies rather than using a full random mechanism in basic HSA (as shown in figure 2) to update the parameter values of harmony memory considering rate and pitch adjusting rate instead of a fixed value. A real world data set on Nurse roster problems collected from a case study hospital in Malaysia was used for performance evaluation. The result showed that EHSA outperforms the Basic Harmony Search Algorithm and the Adaptive Harmony Search in both quality of roaster and computation time.

Chananes, et al. [9] applied a collaborative population based Particle Swarm Optimization (PSO) to the assignment

of home care workers in UK. The provision of community care services and thus the scheduling of home care workers is provided by the local authorities in the UK. PSO is used to optimize the distance travelled by care givers by minimizing the distance to be covered by the care givers while ensuring that the capacity and service time window constraints were maintained. The approach is a systematic improvement of the task of scheduling home care worker by applying a continuous PSO algorithm for the scheduling problem. The proposed method was validated and compared with the conventional methods using real problem sample instances. The proposed method was shown to be a better scheduling solution.

Satoshi et al. [10] proposed a method for the prediction of Emergency Cases (EMS) which involves the correlation of existing EMS with demographic information. A modified group Genetic Algorithm was used to compare current and future optimal locations and numbers of ambulances. Sets of potential locations were evaluated in terms of current and predicted EMS case distances to those locations. The result gotten showed that predicted EMS demands shows there will be an increase of EMS by 2030. The optimal locations of ambulances based on future EMS cases were compared to current locations and with optimal location modeled on current EMS case data. Optimizing the location of ambulance stations location reduced the average response times.

Mohammed et al. [11] investigated parameter settings in Harmony Search Algorithm (HSA) and its application to effectively solve complex Nurse Rostering Problem (NRP). Two different sample sets of rostering cases were used for the performance evaluation of HAS. The first sample set was obtained from a Malaysian hospital and the roster result showed that HSA yielded a better roaster than when GA is used for the rostering problem. The second set is a benchmark nurse rostering problems widely used in the literature. Again, the proposed HSA yielded a better roster results than some of the commonly used meta-heuristic algorithm in the literature.

Herts, and Lahrichi [12] also considered the scheduling problem of assigning nurses to patiences for home care services. The researchers used Tabu Search algorithm to solve a mixed integer programming Model with some non-linear constraints and non-linear objective. Simplification of the work load measure leads to a linear mixed integer program which was optimized using CPLEX.

Margarida and Margarida [13] applied several versions of GA to the nurse rostering problem with hard constraints. The GA associates each individual in population to a pair of chromosome which represents a permutation of tasks and nurses. Those permutated chromosomes served as input to a roaster generating module. The various versions of the GA were tested with real life data collected from a Lisbon hospital. The GAs yielded a good quality roaster.

Riise and Burke [14] presented a model for solving the problem of assigning operation rooms and dates to a set of surgeons, and also assigning each day's surgery to rooms. The daily surgery assignment is done for each surgeon to avoid duplications. The proposed algorithm for the scheduling uses simple relocate and two-exchange neighbourhoods governed by an iterated local search framework. The problem's search space associated is analysed for three typical fitness surfaces, representing different compromises between patient waiting time, surgeon overtime and waiting time for children in the day of surgery. The analysis showed that for the same problem instances, the different objectives gave fitness surfaces with quite different characteristics.

The following figures present some of the algorithms:

```
Do until each physician is assigned to one slot,
    find unassigned physician with largest unassigned patient load
    P1
    find slot with largest available patient load capacity S1
        If S1 > P1
            Assign P1 to S1
            Update patient load capacity and room capacity for
            S1
        ELSE if S1 <P1
                Display informative message and
                    terminate algorithm.

        END IF
Loop
```

Fig. 1: Pseudo code for physician greedy assignment algorithm. [1].

```
Begin
   For (i=1 to HMS) do
     Xi=O
      While coverage demand of night shift not met do;
        allocate two patterns of one-week from the pool of solutions
                that include night shifts cyclically to Xi.

end while
      While not reading the end of Xi do
          allocate two patterns of one-  week from the pool of solutions of
          morning and evening shifts.

 end while
      Calculate the PV of Xi
      add Xi to HM

 end for
    Sort the solutions based on its
          associated penalty value in HM

END
```

Fig. 2: Pseudocode for nurse roistering using EHSA(Masri et al, 2013)

IV. results and discussion

Table 1 presents a summary of the studies reviewed. The table shows the study and the comparative analysis conducted by the authors.

TABLE I. SUMMARY OF REVIEWED LITERATURE

Author	Study	Comparative analysis	Conclusion
Mohammed et al. (2013)	Solving healthcare facility location problem with new heuristic algorithm	The study compared the results obtained in IPSO, GA and Lingo 8 software	Concluded that the IPSO algorithm is faster than the two other algorithms.
Muting and Mboha (2013)	Home Healthcare worker scheduling: A group Genetic algorithm approach	The study Compared The Performance Of Group GA (GGA), PSO, GA, and Parallel PSO (PPSO).	The study concluded that the group genetic algorithm provided a good solution within reasonable computation time compare to other algorithm.
Gerhard et al. (2015)	Metaheuristic for solving a multimodal Home-health care scheduling problem	The study compared the heuristic algorithms i.e. variable neighborhood search, memetic algorithm, scatter search and simulated annealing.	The research was concluded the variable neighborhood search out-performed others in reasonable time and produces valid solution within a few seconds.
Masri et al. (2013)	Enhance Harmony Search Algorithm for Nurse Roaster Problem.	The study compared the performance of Enhanced Basic Harmony Search Algorithm (EHSA), Basic Harmony Search and Adaptive Harmony Search (AHS).	The result of the comparison showed that EHSA produced high quality roasters in short execution time compared to Basic Harmony Search Algorithm and Adaptive Harmony Search.
Chananes et al. (2007)	PSO Based Algorithm for home care worker scheduling in the UK.	The study compare the performance of particle swarm optimization to other exact solution.	The method PSO was implemented, tested and compared to existing exact solutions and was proved to be better solution.
Satoshi et al. (2010)	Using Genetic Algorithm to optimize current and future health planning.	The Research used modified group genetic algorithm to compare current and future optimal locations and number of ambulances.	The modified group Genetic algorithm used in the optimizing the location of ambulance stations, the algorithm reduced the average response time.
Mohammed et al 2013	A harmony search algorithm for nurse roistering problems.	The study compare the performance of Harmony Search Algorithm and Genetic Algorithm on nurse roistering problem.	The comparative study showed that the Harmony Search Algorithm performed better than the Genetic Algorithm.
Herts and Lahrichi 2009	A patient Assignment Algorithm for Home care services.	The researchers used Tabu Search Algorithm to solve a mixed integer programming model.	The work simplified the work load measure.
Margarida and Margarida 2007	A Genetic Algorithm Approach to Nurse Rerostering Problem.	The study developed several versions of Genetic Algorithm which were used for each coding.	The heuristics were tested with real data and good quality solutions were yielded.
Riise and Burke 2011	Local search for surgery admission planning problem.	The research presented an algorithm which uses simple relocate and two-exchange neighborhood governed by an iterated local search.	The analysis showed that for the same problem instances, the different objectives gave fitness surfaces with quite different characteristics.

Most of the reviewed literature compared metaheuristic algorithms and exact scheduling allocation. There is need to include the comparison of heuristic algorithm to be able to ascertain the best solution between the methods.

Space allocation problem is an NP-Hard combinatorial optimization problem which has been attended to by different researchers in the field of Artificial Intelligent (AI) and Operation Research (Ulker and Landa [15]). Due to this combinatorial nature, several methods have been proposed, which include mathematical approach referred to as the exact method, Heuristic method as well as metaheuristic method (Ulker and Landa [16]). Considering the various methods available, metaheuristic is regarded as "one of the best" because of the shortcoming of the exact method that tend to find solution in a solution space without proper attention to the time taking for the solution (Ulker and Landa [16]). Search on schedule/space allocation yields a fair amount of guidelines as described by most researcher reviewed. Unfortunately, most of the health care centres just have guidelines on an automated system because duty allocation in many health centers is mainly a manual process, shouldered on the management to allocate and assign duties.

Medical space is finite in quantity and must be allocated and properly managed to ensure maximum productivity. Where groups with similar interest or who need access to specialized equipment will be grouped together, the success of an healthcare center requires optimal utilization of staff/space assets with the fact that there is a minimum time/space requirements to function optimally.

Scheduling techniques have been widely studied and applied in nursing, physician, patient and time administration. Heuristic and metaheuristic approaches to healthcare center scheduling are attempts to heal worst scheduling process by exchanging a part of a schedule with part from another person's schedule. Basically, this started with an initial solution then developed a neighbourhood search (traditional heuristic or classical heuristic) procedure to find the local optimal. Nevertheless, more recent metaheuristic have been used extensively to solve various scheduling problem ranging from solving health care location problem solving multimodal home-care scheduling problem, optimization of current and future health planning, solving nurse/physicians roaster problem among others.

Metaheuristic approach consists of Tabu Search (TS), Simulated Annealing (SA), Genetic Algorithm (GA), Constraint Programming (CP), Greedy Random Adaptive Search Procedure (GRASP), Ant Colony Optimization (ACO) and Artificial Bee Colony (ABC) among others.

V. CONCLUSION AND RECOMMENDATION

This paper concludes based on the past work reviewed in healthcare schedule that the use of heuristic and metaheuristic approach to different scheduling process in a healthcare center has tremendously reduced the problems that usually emanate from the manual and exact method of scheduling in the past, such as lengthy time required during schedule, staff scheduling at the mercy of officer in charge, human error, over burden of some staff while others will be less burdened etc.

The approach now makes the scheduling and allocation less cumbersome and the entire process can be carried out within some few minutes as against days/months in the other methods. This will help in proper control and management of diseases as the administrative process now involves the use of ICT tool (algorithm) which gives room for proper and fair scheduling. This will encourage proper discharge of duties by the medical personnel.

In view of the above conclusion the use of ICT in automation process of healthcare management. The use of ICT should be encouraged in the management is beneficial of healthcare services as such will facilitate efficient and effective service delivery to patient.

References

[1] Dustin Banet: Heuristic Scheduling for Clinical Physicians. M.Sc thesis submitted to the Department of Industrial Engineering University of Louisville, 2009.

[2] www.un.edu.?

[3] Burke, E.K., Cowling, P. & Landa-Silva. Hybrid population-based metaheristic approaches for the space allocation problem. Proceeding of the 2001 Congress on Evolution Computation (CEC 2001), Vol. no. 2019 page 232-239. 2001

[4] www.healthcapitalinvestmentbranch@Ontario.ca

[5] Mohammed Taghi Valipour, Arman Nedjati and Reze Kazemi: Solving Health Care facility location problem with New Heuristic Algorithm method. International Journal of modeling and optimization Vol. 3, No. 1, page1-10, 2013.

[6] Mutingi, M. and Mbohwa C.: Home Health Care worker scheduling: A Group Genetic Algorithm Approach. Proceeding of the world Congress on Engineering Vol I, WCE, London, U.K 2013.

[7] Gerhard H, Matthias P, Andrea R, Jakob P, Giinther R. Metaheuristics for solving a multimodal home-healthcare scheduling problem, central European Journal of operation research. Springer Verlag, , Vol. 22 (&), pp. 89-113. 2015 (https://hal.inria.fr/hal-01224625).

[8] Masri .A, Mohammed .H, Mohd. Z, and Zulkifli: Enhance Harmony Search Algorithm for Nurse Roaster Problem. A Journal of Applied Sciences 13(6) ISSN 1812-5654 pp 846-853. 2013.

[9] Chananes A, Pisal .Y, and Paul .R, PSO – Based Algorithm for home care worker scheduling in the UK. Journal of Computer & Industrial Engineering Vol 53, issue 4, page 559-583, Nov. 2007.

[10] Satoshi .S, Alexis .J, Hiroshi S and Chris B. using genetic algorithms to optimize current and future health planning. International Journal of Health Geographics Vol. 9:4. Pages 1-10 2010. http://www.ij-health geographic.com/content/9/1/4.

[11] Mohammed Masri .A, Nasser, R, and Roug Q (2013) A harmony search algorithm for nurse rostering problems. Journal of information sciences Vol 233 pages 126 – 140. 2013.

[12] Hertz, A. and Lahrichi, N. (2009) A Patient Assignment Algorithm for Home Care Services. Journal of Operation Research Society Vol. 60, issue 4, pages 481-495. April 2009.

[13] Margarida, M. and Margarida, V.: A Genetic Algorithm Approach to Nurse Rerostering Problem Journal of Computer & Operation Research Vol. 34, issues, pages 667-691, 2007.

[14] Riise, A. and Burke, E.: Local Search for the surgery admission planning problem. Journal of Heuristics. Vol. 17 issue 4 pages 389-414, 2011.

[15] Ulker, D., and Landa, S.: Designing Difficult office space allocation problem instance with mathematical programming, proceedings of 10th International Conference on Experimental Algorithm, page 280-291, 2012.

[16] Ulker, D., and Landa, S.: Evolutionary Local Search for solving the office space allocation problem. Work Congress in Computational Intelligence. Page 3573-3580, 2012.

Data Mining Classification Techniques for the Diagnosis, Treatment and Management of Diabetes Mellitus: A Review

Idowu Dauda OLADIPO
Department of Computer Science,
University of Ilorin, Ilorin, Nigeria
idoladipo@gmail.com

Abdulrauph Olanrewaju BABATUNDE
Department of Computer Science,
University of Ilorin, Ilorin, Nigeria
babatunde2yk@yahoo.com

Abstract: **Diabetes is a universal persistent ailment as a result of variations in regular diets. Many researchers have employed data mining techniques in the prediction and diagnosis of different diseases including diabetes mellitus. The need to review and analyze different methods that has been used in the areas of data mining classification for predicting diabetes mellitus is very important. The paper evaluated some selected studies on classification techniques used in diagnosis and treatment of diabetes mellitus. Findings from this study would go a long way at informative and bridge the recognized gaps for an improvement of the present study and further studies in this subject area.**

Keyword: *Data mining, Diabetes mellitus, Classification, Diagnosis, Prediction.*

I. NTRODUCTION

With the recent advancement in Information Technology (IT), it is a well-known fact that our society is Knowledge driven. Knowledge remains one of the significant assets of any organization [9]. The important role of IT in healthcare system cannot be underestimated. The great demand of IT applications in area of healthcare has provided hope for people to have access to quality and low cost healthcare services. The use of computer system in Hospital Information management Systems (HIMS) has delivered a convenient room to valuable clinical data [2]. Healthcare setting is very rich in information but poor in knowledge. The use of Data mining techniques can provide very useful knowledge if well applied in the healthcare system.

The data mining technique can be seen as an interdisciplinary discipline because it originated from statistics and machine learning algorithm field. The field has advanced from its origin to include database design, artificial intelligence, pattern recognition and visualization. It has advanced from the nascent stage to include pattern recognition, database design, artificial intelligence, visualization. The study of identified data mining was conducted by Massachusetts Institute of Technology (MIT's) in 2001 as one newly discovered emerging technologies that could turn around the world for good if it is well applied [7]. The analysis of large set of data to locate concealed relationship and to summarize the data can be done using data mining methods this will summarize the data into a comprehensible and relevant one [7].

According to [3], the analyzes of data from various perspectives and transforming it into more useful information could be done easily using data mining. Theoretically, data mining is the process of searching for correlations, associations or patterns among a large relational database. Knowledge discovery in database is an important step in data mining technique that is discovering and showing essential data pattern [3]. Data mining which involves the method of obtaining relevant patterns in raw data is also a very useful and powerful tool in Knowledge Discovery in Databases (KDD), data pattern processing, extraction, discovery and harvesting of information [1].

[1] data mining which is also known as knowledge discovery is defined as the non-trivial process of identifying valid, potentially useful, novel and ultimately reasonable designs in data. It enable the generation of scientific theories from large experimental sets of data and biomedical sciences database. The great achievement of it is the application is highly felt in all academics and societal fields like industry, e-business, economic, accounting and others which it led to is application in KDD in other sectors. Among the newly discovered is the healthcare system. Data generated in healthcare are very big, among which are patient's centric data, resource management data and transformed data. Medical sciences and organizations must have ability to analyze data. All information pertaining to patients must be stored and computerized. The use of data mining technique will surely help in supplying answers to many important and serious questions related to healthcare system. The peculiarity in system of healthcare and patient privacy, data mining duty in healthcare domain is not to practice medicine but to assist medical practitioners in obtaining useful information so as to improve the management, diagnosis, treatment and controlling of diseases [9].

The increase of diabetes in the developed countries is growing geometrically in such a way that it can be referred to as

epidemic. The data obtained from International Diabetes Foundation (IDF) in 2005 clearly stated that more than 246 million people in world are receiving treatment for this life intimidating disease. [10]. IDF in their reports gave a probabilistic figure that by the year 2025 over 300 million people worldwide will be confirmed having diabetes mellitus [11]. The real number of individuals with diabetes is higher since by certain epidemiological studies, on each diagnosed person there is one non-diagnosed patient [10]. There is no doubt that there is increase in number of people with diabetes fit to the group of active population. The reduction in impaired fasting glucose (IFG) and fasting plasma glucose (FPG) stages were two things medical doctors used to establish that a person is having diabetes mellitus. This focus of this study to review work using data mining approaches for diagnosis, management and treatment of diabetes mellitus.

II. TECHNIQUES FOR REVIEW OF RELATED STUDIES

The method used in this review work reported in the literature on the data mining and classification techniques used in diagnosis, treatment and management of diabetes mellitus was through the survey of journals and publications. The study also includes books chapters, dissertation, working papers and conference papers. Most recent publications are considered in this appraisal work. Findings form this evaluation would go a long way in coming up with a hearty method to further the research in the prediction for the treatment of diabetes.

III. RELATED WORK

Many studies have been conducted past for the prediction and diagnosis and treatment of diabetes mellitus. Many researchers have applied different data mining techniques for predicting diabetes:

Decision tree, logistic regression and Artificial neural networks were used for prediction of diabetes among two sets of people from china. Designed classification accuracy achieved is differed respectively, 73.23% for logistic regression, 77.8% for Decision tree, 77.8% for Artificial Neural Network. The experimental result showed that Decision tree outperformed logistic regression and ANN [21].

The disease is classified into two based on what causes the diabetes: firstly, DNA clarified level was based on genetic susceptibility and another one that is associated with other diseases or conditions [34]. The Swedish National Diabetes Register (NDR) received complains about smoking habits that they should study the relationship between smoking glycemic control and micro albuminuria by identifying trends associated with smoking habits. They deduced that smoking habits in patients with this ailment were of increase mostly among the young females with type 1 diabetes, also the middle-aged type 1 and 2 diabetes patients were of increase. The study recommended smoking cessation campaigns for these categories of people. The habits were associated with poor glycemic control and micro-albuminuria that were not part of the characteristics of other study [35].

The weight loss was investigated among people enrolled in an adapted Diabetes Prevention Program (DDP). The people were able to monitor their dietary intake and increase in physical activity marked regularly. From the research, the importance of supporting lifestyle interventions to initiate and maintain dietary self-monitoring and increase levels of their physical activity were very necessary [36]. Another study was performed to check the prediction analysis for treatment of hypertension using regression based system a branch of data mining techniques [37]. Another study was conducted using logistic regression model in drug treatment for diabetic people, the study concluded that the drug treatment used for the young group have side effects and patients of old age need drug administration immediately diabetes is discovered in them since no other measure is available to be used [6].

A study conducted using fuzzy ontology model in studying the diabetes knowledge with relation based structure for fuzzy diabetes ontology. The work was known for its effective decision support model [22]. KNN have been used in it modified forms by many researchers to improved existing model for efficiency. Pima Indian diabetes dataset used in testing the efficiency of a class wise K-nearest algorithm. The testing data has been classified into class label corresponding to the lowest distance, the model recorded a 78.16% accuracy for C-kNN [23].

A study was performed by Xue-Hui, et al in 2013 to compare three data mining algorithms namely: artificial neural networks, decision tree models and logistic regression for their performance and effectiveness for predicting diabetes and pre-diabetes using common risk factors. Two communities in Guangzhou, in China was used as an experiment; the use of questionnaire was used to obtain relevant information on family diabetes history, lifestyle risk factors, anthropometric measurements and their demographic characteristics was among data collected from the participants. Among the people scrutinized 735 patients was confirmed having diabetes and 752 control were engaged. The authors designed three predictive models based on the questionnaire administered with 12 input variables with one output variable. The models were evaluated based on their accuracy, sensitivity and specificity [4].

The results of the three models were evaluated and the following results were recorded:

Table 1: Results of the three models

[17] Models	[18] Classification accuracy	[19] Sensitivity	[20] specificity
[21] Logistic regression	[22] 76.13%	[23] 79.59%	[24] 72.74%
[25] Artificial Neural networks	[26] 73.23%	[27] 82.18%	[28] 64.49%
[29] The decision tree	[30] 77.87%	[31] 80.68%	[32] 75.13%

C5.0 a decision tree model outperformed the other two model in term of accuracy, ANN performed best in term of sensitivity and C5.0 better than others in term of it specificity.

Chao-Ton et al, 2006 designed a prediction model for diabetes type 2 patients using scanning data obtained from anthropometrical body surface. They made used of four

different of data mining techniques namely: rough set, logistic regression, decision and neural network in selecting relevant symptoms to predict diabetes. The accuracy of these models was evaluated. Amount of trunk, waist circumference, volume of right leg, right thigh circumference, and age were some of the symptoms of diabetes [5]. Accuracy of decision tree and rough set out perform logistic regression and neural network, based on the rules extracted from the anthropometrical data by using decision tree. The result obtained from the study is useful in medicine for diagnosis of diabetes and also very useful for diabetes preventive measure [5].

[6] used regression based technique for analysis of diabetic treatment, oracle Data Miner (ODM) was used as a mining tool in determining treatment of diabetes. Support vector machine algorithm was employed the experimental evaluation. The dataset of Non-communicable Diseases (NCD) of Saudi Arabia were obtained from World Health Organization (WHO) for data analysis. The dataset was used to evaluate the effectiveness of the treatment of different age groups. Five age group that were obtained was merged into two age group, denoted as p(y) and p(o) that stand for young and old age group respectively. This dataset was used to find the preferential orders of treatment of the merged age group. It was concluded that old age group should be administered drug immediately with other treatment while the young group can be delayed in order to prevent the side effects the treatment may cause since there are no any another measure available.

[8] proposed and developed a data mining model for which will predict a suitable dosage planning for diabetes patients. Medical records of 89 different patient were used in this study. 318 diabetes assays were extracted using these patient records. ANFIS and Rough Set methods were used for dosage planning objective. According to the results of ANFIS and Rough Set methods, ANFIS is a more successful and reliable method for diabetes drug planning objective when compared to Rough Set method.

Another author evaluated the potential use of classification methods in diabetes the data mining methods such as ANN, rule based and decision tree was used in the proposed evaluation. A diabetic patient medical dataset was used in classification techniques in determining their evaluations [9].

Aiswarya, Jeyalatha & Ronak [12] used decision tree and Naïve Bayes data mining methods to find solutions to diagnosis problems in diabetes patients by analyzing the patterns of dataset obtained through classification techniques. The paper proposed a quicker and more efficient technique of diagnosing the disease, leading to timely treatment of patients [12].

[13] the authors used Naive Bayes, RBF Network, and J48 data mining algorithms to diagnose type II diabetes. The algorithms performed diagnosis using Weka. Also, the algorithms were compared to determine which one was more accurate in diagnosis of type II diabetes. The results revealed that Naive Bayes, having accuracy rate of 76.95%, enjoyed the highest accuracy for diagnosis of type II diabetes.

Study conducted by Ananthapadmanaban & Parthiban in 2014 using Naïve Bayes with Support Vector Machine (SVM) in finding early detection of eye ailment in diabetic retinopathy. The paper test the accuracy, sensitivity and specificity of the two data mining classification methods, from the study it was discovered that Naïve Bayes outperformed SVM in term of accuracy with 83.37%. The system shows that data mining can help in retrieving useful correlation from variables that are not direct indicators to the selected predicting indicators [14].

Velide & Lakshmi, 2014 using different type of data mining techniques using accuracy and time to evaluate the performance of Naiva Bayes, ANN, decision trees, fuzzy logic, KNN, J48 (C4.5) JRip and Genetic Algorithms. The paper selected nine (9) attributes to evaluate the selected data mining algorithms used. The selected attributes are as follows: WerSex, Diabetes probability, 2 hours Serum Insulin, number of times Pregnant, Diabetes Pedigree type, BMI, Skin fold thick, Plasma glucose, and Diastolic B.P. It was recorded that J48 (C4.5) simple are efficient classifier diabetes data compare to others data mining methods [15].

A research conducted by Joseph, Colin, & Peter in 2002, using data mining approach by examining one dataset warehouse. They used data issues, analysis problems to obtain results, the dataset of New Orleans diabetic warehouse were used to obtain the data that were used. it is a very big integrated health system with 30,383 diabetic patients. This was used to translate a complex relational database with time series and sequencing information to flat file that were in data mining challenges. They measured two different variables, comorbidity index with HgbA1c and a measure of glycemic control were related with their outcomes.

Decision tree and regression tree (CART®) with binary target variable of HgbA1c>9.5 and 10 predictors were used. They are: age, sex, emergency department visit, end stage renal disease, retinopathy, cardiovascular disease, hypertension, dyslipidemia, comorbidity index and office visits, young age shows important variable that is associated with bad glycemic control and not comorbidity index and patients have related diseases as expected. The target diabetics with bad HgbA1c values is high in those age that are less than 65 years old than people that are older is 3.2 times more. The use of data mining is a novel method of discovering association that very useful in clinicians and administrators of treatment and drugs for patients [16].

In United States, it was reported that the direct and indirect cost of diabetes diseases is $132 billion [3,5]. Hundreds of Millions of people around the world are suffering from diabetes disease and in USA alone the population of people suffering from this ailment is about 15.7 million [17,18]. It have been proved that there is no method that is capable of eradicating diabetes, medical sciences are trying to push for ways of fighting the ailment.

The causes of diabetes in patients have been traced to the following: lifestyle, age, inheritance related ailment and so forth in finding the root cause of diabetes [17]. The many prevalence number of factors and historical dataset can be

used to generate inference rule on the causes and effects of diabetes using data mining tools, these has well helped in knowledge discovery. The authors discussed different steps that are used in mining diabetes data, and using a case study of how data mining can be carried out in detection and diagnosis of diabetes using Hong Kong, USA, Poland and Singapore are their case study [17].

Many researchers have tried to use an expert system in diagnosis of diabetes there is still some setback since patient still undergo various medical tests used has input variables for computerized diagnostic system. The cost of diagnosis still remains almost the same, Kumari & Singh, 2013 designed a system based on ANN and claimed that the system model allows users themselves diagnose whether they suffering from diabetes mellitus or not. All the patient need is to provide some physical parameter values based on the information's provided the system will detect whether the patient is having the ailment or not. MATLAB was used to design the proposed expert system with the used of ANN algorithms [18].

In Lingaraj, Devadass, Gopi, & Palanisamy, data mining methods are used for an expert system for diabetes diagnosis and treatment. The ailment is a family of metabolic disease that are high blood sugar levels for long period of time. Authors do much of literature survey that are related to various data mining methods in diabetes prediction. There are still ongoing research work on the presentation of diabetes in medical sciences. Many authors have studied the causes of diabetes using data mining approaches, only very few clinical risk factors are used for the mining. Many important variables are neglected in their analysis especially the pre-diabetes health conditions. The results by such methods may not shows the relevant risk factors and pattern recognition of diabetes correctly [19].

For the above reasons Akkarapol and Jongsawas, categorize the analysis into three areas based on patients healthcare costs. The authors then examine whether more complex analytical models techniques using different data mining in SAS®, Enterprise MinerTM 7.1 can be used to predict and explain the causes of increasing diabetes in patients categorized in each cost analysis. The analysis shows that the causes of diabetes in adult patients are as follows: heart attack, marital status, dental checkup, high blood pressure, age, cholesterol, adult BMI, total income, and sex. These are some of the key risk factors discovered by their models [24].

Diabetes has been identified to be a modern society disease, there are also huge amounts of medical data that currently available and useful knowledge after applying powerful data analysis tools. Using a reliable prediction method to diagnose diabetes will support medical sciences in this field [25]. The research on using prediction analysis for diabetes diagnosis using ANN as a data mining techniques. The WEKA software was employed as data mining tool for diabetes diagnosing. The authors used the Pima Indian diabetes database from UCI server for their analysis. The dataset was studies and analyzed to build an effective model used in predicting and diagnosing the ailment [25].

Data mining have been proved to be a diversity of methods in investigate large data keeping in mind the end result to find hidden knowledge. The authors used a descriptive data mining approach and to devise association standards to envisage diabetes behaviour in arrangement with particular life style parameters, which include physical activity and emotional states, mostly in elderly diabetics. The proposed models are based on Random Forest Classifier [26].

IV. CONCLUSION

The use of data mining have given an idea and revolution to process and investigate with immense measures of data into useful and helpful information for decision making that is an essential piece of medical service management [26]. Diabetes is special amongst the most well-known non transmittable diseases in the world. It is considered to be the fifth cause for death in most of the nation [38]. These diabetics can lead a typical life through mix of a day by day insulin treatment, solid eating regimen, close checking and normal physical activity [39]. The works review reveals many results on diabetes with the use of data mining. These will help the researchers to be able to work further on diabetes using data mining and other artificial intelligence methods in diagnosis, treatment and management of diabetes mellitus. From all the appraisal presented above, some of the past work do not really discuss the management of Diabetes Mellitus. Future work should therefore be on the development of an efficient technique to improve on diagnosis and management of Diabetes Mellitus. The future research methodology should involve other artificial intelligence using fuzzy expert system for the management of diabetes mellitus to get an efficient result.

REFERENCE

[1]. Andrea LH, Hsinchun C, Susan MH, Bruce RS, Tobun DN, Robin RS, Kristin MT (1999) Medical data mining on the internet: research on a cancer information system. Artif Intell Rev 13:437–466.

[2]. Wasan, Siri Krishan, Vasudha Bhatnagar, and Harleen Kaur. The impact of data mining techniques on medical diagnostics, Data Science Journal 5.19 2006, 119-124.

[3]. Jacob, S., and Ramani, G., Design and Implementation of a Clinical Data Classifier: A Supervised learning approach, Res. J. Biotech, Vol.8, No.2, pp.16-26, 2011

[4]. Xue-Hui Meng, Yi-Xiang Huang, Dong-Ping Rao, Qiu Zhang, and Qing Liu, Comparison of three data mining models for predicting diabetes or prediabetes by risk factors, Kaohsiung Journal of Medical Sciences 29, Pp. 93-99, 2013.

[5]. Chao-Ton Su, Chien-Hsin Yang, Kuang-Hung Hsu, Wen-Ko Chiu, Data Mining for the Diagnosis of Type II Diabetes from Three-Dimensional Body Surface, Anthropometrical Scanning Data, Computers and Mathematics with Applications, 2006, 51, Pp. 1075-1092

[6]. Abdullah A. A., Mohammed G.A., Mohammad K.S. Application of data mining: Diabetes health care in young and old patients, Journal of King Saud University – Computer and Information Sciences, 2013, 25, Pp. 127–136.

[7]. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang J. F, Hua Library, Data mining in healthcare and biomedicine:a survey of the literature. Journal of medical systems, 2012, 36(4):2431–2448.

[8]. Emirhan G.Y., Adem K., Tamer U., Dosage planning for diabetes patients using data mining methods, Procedia Computer Science 2011, 3, Pp.1374–1380.

[9]. Harleen, K, & Siri, K.W. Empirical Study on Applications of Data Mining Techniques in Healthcare, Journal of Computer Science 2 (2): Pp. 194-200, 2012.

[10]. International Diabetes Federation. Diabetes Atlas. 3rd edn. Brussels: International Diabetes Federation; 2006.
[11]. Danijela TADIĆ, Predrag POPOVIĆ & Aleksandar ĐUKIĆ, A Fuzzy Approach to Evaluation and Management of Therapeutic Procedure in Diabetes Mellitus Treatment, Yugoslav Journal of Operations Research Vol. 20, Number 1, 99-116, 2010.
[12]. Aiswarya Iyer, S. Jeyalatha and Ronak Sumbaly, Diagnosis of Diabetes Using Classification Mining Techniques, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.1, Pp. 1-14, 2015
[13]. Sadri .S., Amanj .M., Ramin .H., Zahra .P. & Kamal .C., Comparison of Data Mining Algorithms in the Diagnosis of Type II Diabetes, International Journal on Computational Science & Applications (IJCSA) Vol.5, No.5, Pp. 1-12. 2015.
[14]. K. R. Ananthapadmanaban & G. Parthiban, Prediction of Chances - Diabetic Retinopathy using Data Mining Classification Techniques, Indian Journal of Science and Technology, *Vol 7(10), 1498–1503,* 2014.
[15]. VelidePhani K. & Lakshmi V., A Data Mining Approach for Prediction and Treatment of Diabetes Disease, International Journal of Science Inventions Today, 3(1), Pp. 073-079., 2014
[16]. Joseph L.B., Colin R.G., & Peter J.F., Data Mining a Diabetic Data Warehouse, *Artificial Intelligence in Medicine*, Pp. 1-39, 2002.
[17]. Indranil Bose, Data Mining in Diabetes Diagnosis and Detection, Encyclopedia of Data Warehousing and Mining, Pp 257-261, 2005
[18]. S. Kumari and A. Singh, A data mining approach for the diagnosis of diabetes mellitus, *Intelligent Systems and Control (ISCO), 7th International Conference on*, Coimbatore, Tamil Nadu, India, Pp. 373-375, 2013.
[19]. Lingaraj, H., Devadass, R., Gopi, V., Palanisamy, K. Prediction of Diabetes Mellitus using Data Mining Techniques: A Review. Journal of Bioinformatics & Cheminformatics, 1(1), Pp 1-3, 2015.
[20]. S.Vijiyarani and S.Sudha, Disease Prediction in Data Mining Technique - A Survey, International Journal of Computer Applications and Information Technology. 2(1), 2013.
[21]. X. H. Meng, Y. X. Huang, D. P. Rao, Q. Zhang and Q. Liu., Comparison of three data mining models for predicting diabetes or prediabetes by risk factors. Kaohsiung Journal of Medical Sciences, 29(2): 93-99, 2013.
[22]. Chang-Shing Lee, A Fuzzy Expert System for Diabetes Decision Support Application. IEEE transactions on systems, man, and cybernetics - part b: cybernetics, 2011 41(1).
[23]. Y. Angeline Christobel and P.Sivaprakasam, A New Classwise K Nearest Neighbor (CKNN) Method for the Classification of Diabetes Dataset. International Journal of Engineering and Advanced Technology (IJEAT), 2(3), 2013.
[24]. Akkarapol .S. & Jongsawas .C., An Analysis of Diabetes Risk Factors Using Data Mining Approach, 2012.
[25]. Arwa .A., Maram .A., Nasebih .A. & Mohammed A.A., Using Prediction Methods in Data Mining for Diabetes Diagnosis, 2014.
[26]. Mani Butwall & Shraddha Kumar, A Data Mining Approach for the Diagnosis of Diabetes Mellitus using Random Forest Classifier, International Journal of Computer Applications, Volume 120 – No.8, Pp. 36-39, 2015.
[27]. Al-Nozha, M.M., Al-Maatouq, M.A., Al-Mazrou, Y.Y., Al-Harthi, S.S., Arafah, M.R., Khalil, M.Z., Khan, N.B., Al-Khadra, A., Al-Marzouki, K., Nouh, M.S., Abdullah, M., Attas, O., Al-Shahid, M.S., Al-Mobeireek, A., Diabetes mellitus in Saudi Arabia. Saudi Medical Journal 25 (11), 1603–1610, 2004.
[28]. Breault, Joseph L., Goodall, Colin R., Fose, Peter J., Data mining a diabetic data warehouse. Artificial Intelligence in Medicine 2002, 26, 37–54.
[29]. Charpentier, G., Gene` s, N., Vaur, L., Amar, J., Clerson, P., Cambou, J.P., Gue´ ret, P., Control of diabetes and cardiovascular risk factors in patients with type 2 diabetes:a nationwide French survey. Diabetes & Metabolism 2003, 29, 152–158,
[30]. Sigurdardottir, Arun K., Jonsdottir, Helga, Benediktsson, Rafn, Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis. Patient Education and Counseling 67, 21–31, 2007.
[31]. Yokotaa, Ichiro, Amemiya, Shin, Kida, Kaichi, Sasaki, Nozomu, Matsuura, Nobuo, Past 10-year status of insulin therapy for preschool-age Japanese children with type 1 diabetes. Diabetes Research and Clinical Practice, 2005, 67, 227–233.
[32]. Craiga, Maria E., Jonese, Timothy W., Silinka, Martin, Pingg, Yeo Jing, Diabetes care, glycemic control, and complications in children with type 1 diabetes from Asia and the Western Pacific Region. Journal of Diabetes and its Complications 2007, 21, 280–287.
[33]. Soriguer-Escofet, F., I. Esteva, G. Rojo-Martinez, S.R. de Adana and M. Catala et al., Prevalence of Latent Autoimmune Diabetes of Adults (LADA) in Southern Spain. Diabetes Res. Clin. Pract., 2002, 56: 213-220.
[34]. Richards, G., Rayward-Smith, V.J., Sonksen, P.H., Carey, S., Weng, C., Data mining for indicators of early mortality in a database of clinical records. Artificial Intelligence in Medicine, 2001, 22,215–231.
[35]. Nilsson, P.M. et al, Smoking is associated with increased HbA1c values and microalbuminuria in patients with diabetes—data from the National Diabetes Register in Sweden. Diabetes & Metabolism, 2004, 30, 261–2688,
[36]. Harwell, Todd S. et al, Factors associated with achieving a weight loss goal among participants in an adapted Diabetes Prevention Program. Primary Care Diabetes, 2010.
[37]. Almazyad, A.S., Ahamad, M.G., Siddiqui, M.K., Almazyad, A.S., Effective hypertensive treatment using data mining in Saudi Arabia. Journal of Clinical Monitoring and Computing 24 (6), 391–401, 2010.
[38]. IDF Diabetes Atlas, 6th edition, "International Diabetes Federation", 2013, Online available at: http://www.idf.org/diabetesatlas
[39]. "Definition, classification and diagnosis of diabetes, Prediabetes and metabolic syndrome", Canadian Journal of Diabetes, Vol. 37, Canadian Diabetes Association. 2013. Online available at: www.canadianjournalofdiabetes.com

Redundant Residue Number System Based Fault Tolerant Architecture for Communication System

Kadri Akeem Femi
Department of Computer Science
Kwara State University,malete
Ilorin, Nigeria
akeem.kadri@yahoo.com

Saheed Yakub Kayode
Department of Physical Science
Al-Hikmah University,Ilorin
Ilorin,Nigeria
yksaheed@alhikmah.edu.ng

Gbolagade Kazeem Alagbe
Department of Computer Science
Kwara State University, malete
Ilorin, Nigeria
kazeem.gbolagade@kwasu.edu.ng

***Abstract*-This paper proposed the Redundant Residue Number System based Code Division Multiple Access over a communication channel. The Coded transmission technique applied in a multipath environment has a Bit Error Rate comparable to that of a narrow band radio channel due to the fact that the fading of each subcarrier is frequency non-selective. The existing scheme requires large block of data and higher overhead. Moreover, simple error detection such as parity check bit is too weak for communication in which quality of radio channel is often poor and burst error often occurs. These aforementioned shortcomings reduce the performance of the CDMA. In this paper, an alternative scheme to detect and correct error, in order to increase performance and also to eliminate the higher overhead in the present CDMA scheme is presented. We use length five (5) moduli set [2^n, $2^n - 1$, $2^n + 1$, $2^n - 2^{(n+1)/2} + 1$, $2^n + 2^{(n+1)/2} + 1$], where ($2^n - 2^{(n+1)/2} + 1$, $2^n + 2^{(n+1)/2} + 1$) is the redundant moduli set which is used for the correction of error. The proposed scheme increases the performance of CDMA and provides more capability for fault-tolerance than those similar of the state-of the-art.**

Keywords-Redundant Residue Number System; Code Division Multiple Access; Moduli Set; Forward Converter.

I. INTRODUCTION

The multiple-access technique used in communication system is one way of efficiently allocating a rare communication resource, namely, the radio spectrum. This technique becomes meaningful when a large number of users seek to communicate with each other simultaneously. This sharing of the frequency spectrum must be done in such a way that it does not negatively affect the quality of performance of the system. It is also often desirable to allow a user to send and receive information simultaneously.

This can be achieved by duplexing in the time domain or in the frequency domain [1]. Multiple-access techniques allow several users to communicate simultaneously while providing each user with duplexing capability.

There are three major access techniques that are used to share available bandwidth: frequency-division multiple access (FDMA), time-division multiple access (TDMA), and code-division multiple access (CDMA) These three techniques can be broadly categorized as narrowband and wideband [1].

One of the main advantages of residue number systems is that they facilitate the detection and correction of errors. This arises from the fact that in the residue representation of a number, all the digits are independent; therefore, an error in one digit-position does not corrupt any other digit-position. So if an error occurs in some digit-position, computations may still proceed, through the exclusion of the faulty digit position (and corresponding modulus), provided that either the resulting smaller dynamic range is acceptable, or that the original system had some extra moduli that provide a larger range than that nominally required for the computations to be carried out. Note, though, that while fail-soft capability exists for all operations, error-isolation is not possible with operations that require interaction between all digits of an RNS representation; that is, operations such as division, magnitude-comparison, and reverse conversion.

Types of bit error in wireless communication can be classified as either random errors or burst errors. The random bit error can appear at any location in a sequence of data transmission or radio packet.

II.PROPOSED METHODOLOGY

RRNS QC ENCODER/DECODER

In this section, we explain an efficient structure for the RRNS QC encoder/decoder as shown in the diagram below. It consists of (m+2r) B/R converters, (m+2r) binary sub encoders, (m+2r) binary sub decoders and an R/B converter with error-correcting capability [7].

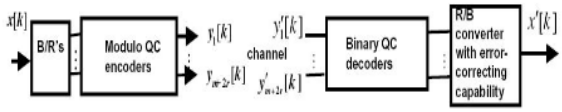


Fig.1 The Encoder/Decoder Diagram [8].

Based on the redundant moduli$\{P_1,P_2,\ldots,P_m, \ldots,P_m+2_r\}$, the B/R converters convert the binary message signal X[k] into its residue format, $(X_1[k], X_2[k], \ldots ,X_m[k],\ldots,X_m+2_r[k])$, where $Xi[k]= \lfloor X[k] \rfloor$ and Pi, i=1, ..., m, ..., m+2r. Then, for each Xi[k], one binary encoder is used to generate the corresponding encoded output Yi[k]. The (m+2r) encoded signals Yi[k] are transmitted through the transmission channel. At the receiver, the binary decoders, which are the inverse of the corresponding encoders, recover each residue message X'i[k]. Finally, the R/B converter with error-correcting capability will detect and correct the r errors and generate the final output message signal X'[k]. The proposed architecture has the following features:

• Binary encoders and decoders are used. Thus, a number of modulo operations required by the design in [3-4] are removed.

• The proposed architecture produces a security in the transmission (QC properties), and has an error correcting capability at the back end of the receiver. Thus, the message received from the final stage is error-corrected.

• If the word length of the input message is *L* bits, the word length of each binary encoder or decoder is approximately log (*L*) bits. Thus, the speed of the encoding and decoding can be very high.

III CONVERSION ALGORITHMS

The Conversion algorithms are;

i. Chinese Reminder Theorem CRT

ii. Forward conversions

ii. Reverse converters

Before we talk about conversion, let's see the reason beyond special set of moduli $\{2^n+1, 2^n, 2^n-1\}$. Let's consider the computation of the residue of integer X with respect to a modulus m. Since X may be represented as an n-bit binary number (xn-1, x_{n-2} … x_0), and its residue with respect to m may be expressed as:

$$|X|_m = |x_{n-1}, x_{n-2} + \ldots + x_0|_m$$
$$|X|_m = |2^{n-1} * x_{n-1} + 2^{n-2} * x_{n-2} + \ldots + x_0|_m$$

From the properties of residues, we have:

$$|X|_m = \left||2^{n-1} * x_{n-1}|_m + |2^{n-2} * x_{n-2}|_m + \ldots + |x_0|_m\right|_m$$

[9].

IV. DATA CONVERSION IN RNS

In any RNS design, input operands are given in either binary or decimal format and must be converted to RNS notation before computation can be performed. The final results must be re-presented in the same way as the input operands. Hence, the need for a binary to residue, residue to binary converter for a successful RNS designs. There are two types of conversions in RNS processor: a forward conversion and a reverse conversion, with the latter one of the major if not the major problem in the full adoption of RNS due to its complex nature [6].

V. RESULT PRESENTATION

A. FORWARD CONVERSION

Using the moduli set

$[2^n, 2^n – 1, 2^n + 1, 2^n - 2^{(n+1)/2} + 1, 2^n +2^{(n+1)/2} + 1]$ [2].

Example

Let us consider the binary number 1010 to RNS using the moduli set above

Let n = 3

= (8, 7, 9, 5, 13)

Let

$(2^n, 2^n – 1, 2^n + 1)$ be our moduli set

Let n = 3 for the above moduli set

(8, 7, 9)

M = 8 × 7 × 9 = 504

Using the binary

1010 RNS (8, 7, 9)

1010 mod 8 will be 2

1010 mod 7 will be 3

1010 mod 9 will be 1

Let

$(2^n - 2^{(n+1)/2} + 1, 2^n +2^{(n+1)/2} + 1)$ be our redundant moduli set

Let n = 3 for the above moduli set

= (5, 13)

Using the binary

1010 RNS (5, 13)

1010 mod 5 will be 0

1010 mod 13 will be 10

We now have (2, 3, 1, 0, 10) RNS (8, 7, 9, 5, 13)

Let

X_1, X_2, X_3, X_4, X_5 be (2, 3, 1, 0, 10) moduli set/redundant

M_1, M_2, M_3, M_4, M_5 be (8, 7, 9, 5, 13) data residue/redundant

Suppose there is error in the RNS representation due to transmission or processing for instance in the following; Example: x_1 is change from 2 to 3 and then the received RNS representation becomes (3, 3, 1, 0, 10). Upon following the general approach of the Chinese reminder theorem (CRT) and using the first (1st) three (3) residue digit and their moduli, we obtain CRT formula .

For the equation on CRT to be used, it means that the moduli set used must be pairwise relatively prime although there has been serious work on moduli sets with common factors [5].

$$X = \left| \sum_{i=1}^{n} m_i |x_i m_i^{-1}|_{m_i} \right|_M$$

$X = | m_1|m_1^{-1}|m_1x_1 + | m_2|m_2^{-2}|m_2x_2 + | m_3|m_3^{-3}|m_3x_3 + \ldots + | m_n|m_n^{-n}|m_nx_n |_M$ [5].

B. REVERSE CONVERSION

Now (x_1, x_2, x_3) RNS (8, 7, 9)

(3, 3, 1) RNS (8, 7, 9)

$M_1 = 63$

$M_2 = 72$

$M_3 = 56$

$M_1^{-1} = | 63^{-1} | = 7$

$M_2^{-1} = | 72^{-1} | = 4$

$M_3^{-1} = | 56^{-1}| = 5$

$M = 8 \times 7 \times 9 = 504$

$X = | 63\times7\times3 + 72\times4\times3 + 56\times5\times1 |_{504}$

$= | 1323 + 864 + 280 |_{504}$

$= | 2467 |_{504}$

$= 451$

$| 451 |_5 = 1$

$| 451 |_{13} = 9$

$| 0 - 1 |_5 = 4$

$| 10 - 9 |_{13} = 1$

ERROR DETECT

Since there is an error in (x_1, x_2, x_3) RNS (8, 7, 9)

We now check for (x_2, x_3, x_4) RNS (7, 9, 5)

CORRECT/RESTORE ERROR

(3, 1, 0) RNS (7, 9, 5)

$M_2 = 45$

$M_3 = 35$

$M_4 = 63$

$M_2^{-1} = | 45^{-1} | = 5$

$M_3^{-1} = | 35^{-1} | = 8$

$M_4^{-1} = | 63^{-1}| = 2$

$M = 7 \times 9 \times 5 = 315$

CONSISTENCE CHECKING

$X = | 45\times5\times3 + 35\times8\times1 + 63\times2\times0 |_{315}$

$= | 675 + 280 + 0 |_{315}$

$= | 955 |_{315}$

$= 10$

$| 10 |_5 = 0$

$| 10 |_{13} = 10$

$| 0 - 0 |_5 = 0$

$| 10 - 10 |_{13} = 0$

Since there is an error in (x_2, x_3, x_4) RNS (7, 9, 5)

We now correct the error in x_1 by $| 10 |_8 = 2$

VI. RESULT

In Table 1 shown RRNS code words of some typical decimal integer messages X in the RRNS with moduli $m_1 = 4$; $m_2 = 5$; $m_3 = 7$; $m_4 = 9$; $m_5 = 11$; $m_6 = 13$ and $m_7 = 17$; where $r_i = X \pmod{mi}$ and $M = 4\times5\times7 = 140$.

TABLE I. SHOWN RRNS CODE WORDS OF SOME TYPICAL DECIMAL INTEGER MESSAGES

Decimal Message X	No redundant Residue digits			Redundant Residue digits			
	r_1	r_2	r_3	r_4	r_5	r_6	r_7
$X_0 = 0$	0	0	0	0	0	0	0
$X_1 = 1$	1	1	1	1	1	1	1
$X_2 = 2$	2	2	2	2	2	2	2
$X_3 = 5$	1	0	5	5	5	5	5
$X_4 = 10$	2	0	3	1	10	10	10
$X_5 = 20$	0	0	6	2	9	7	3
$X_6 = 50$	2	0	1	5	6	11	16
$X_7 = 100$	0	0	2	1	1	9	15

In Table 2 shown RRNS code words of some typical binary messages X in the RRNS with moduli $m_1=4$; $m_2=5$; $m_3=7$; $m_4=9$; $m_5=11$; $m_6=13$ and $m_7=17$; where $r_i = X \pmod{mi}$ and $M= 4 \times 5 \times 7 =140$.

TABLE II. SHOWING RRNS CODE WORDS OF SOME TYPICAL BINARY INTEGER MESSAGES

Binary Message X	No redundant Residue digits			Redundant Residue digits			
	r_1	r_2	r_3	r_4	r_5	r_6	r_7
$X_0 = 0000000$	0	0	0	0	0	0	0
$X_1 = 0000001$	1	1	1	1	1	1	1
$X_2 = 0000010$	2	2	2	2	2	2	2
$X_3 = 0000100$	0	4	4	4	4	4	4
$X_4 = 0001000$	0	3	1	8	8	8	8
$X_5 = 0010000$	0	1	2	7	5	3	16
$X_6 = 0100000$	0	2	3	5	10	6	15
$X_7 = 1000000$	0	4	1	1	9	12	13

CONCLUSION

In this paper, we look at the problems in communication system such as: errors in transmitted data. Then we studied some of existing capabilities in the RNS. These capabilities can be used to improve the existing problems in communication system. The merit of this work is that it reduces traffic rate in communication system. Additionally, RNS has the ability to detect and correct errors in data transmitted while exploiting minimum redundancy.

REFERENCES

[1] A. Omondi and B. Prekumar, "Residue Number System: Theory and Implementation" (2007).

[2] Ahmed A. Hiasat, "VLSI Implementation of New Arithmetic Residue to Binary Decoder".

[3] Wei Wang, Xiaolin Zhang, Chenyang Yang, M. N. S. Swamy and M. O. Ahmad,"RRNS Quasichaotic Coding And Its FPGA Implementation" 0-7695-2294-7/05.2005 IEEE.

[4] H. How, T. H. Liew, E. Kuan, L. L. Yang, and L. Hanzo, "A redundant residue number system coded burst-by-burst adaptive joint-detection based CDMA speech transceiver," IEEE Trans. Veh. Technol., vol. 55, no. 1, pp. 387-397, Jan. 2006.

[5] Gbolagade Alagbe Kazeem & Sorin Dan Cotofona, "A reverse converter for the moduli set" IEEE, 2009.

[6] Kazeem Alagbe Gbolagade"Effective reverse conversion in RNS proccesors. *Phd thesis 2010.*

[7] L. L. Yang and L. Hanzo,"Minimum-distance decoding of redundant residue number system codes," in Proc. IEEE ICC 2001, pp. 2975-2979.

[8] M. Roshanzadeh and S. Saqaeeyan, "Error Detection and Correction in Wireless Sensor Network".I.J. Computer Network and Information Security, 2012, 2, 29-35.

[9] Ahmad q. Al badawi, Ahmad m. Baset & Qasem s. Abu Al-haija, a novel digital adder design based on residue number system with special set of moduli $\{2^n+1, 2^n, 2^n-1\}$.

A Modified Huffman Algorithm and Its Application

Adebayo Kayode Adeleye
Computer science department
University of Ilorin.
Kwara state. Nigeria.
Email. leye.kayode@gmail.com

'Dele Oluwade
Computer Science Department
University of Ilorin.
Kwara State. Nigeria
Email: deleoluwade@yahoo.com

Abstract: **Data compression is expedient for space conservation and bandwidth management in this information age. Therefore, compression of natural language, Yoruba in particular should not be abandoned. With the recent increase in the population of the Yoruba language in the world which is said to be above 60 million people, the increase in Yoruba content on the internet and other digital media is on a rapid increase and so should not be carelessly abandoned: This research work thus focuses on the compression of Yoruba text for the increased population. Huffman algorithm was adopted in this research and subsequently improved to suit data compression of Yoruba text. The basic idea is that instead of storing each character in a file as an 8-bit ASCII or EBCDIC value, shorter codes are then assigned to symbols that occur more frequently and longer codes to those that occur less frequently. By constructing a Full Binary Tree Representation (FBTR), the tree have edges, each of which has a value 0 or 1,('0' is the left child, '1'is the right child). The modified algorithm manipulates each character (alphabet, space, digraph, diacritical marks) in the message file to reduce the size with no data loss even after decompression. The digraph "GB" is seen as a single but not separate alphabet, the upper and lower case of the text are uniquely defined. In view of these, character frequency table is generated based on the number of each character in the text; this frequency is subsequently used to construct the Binary tree that produce the codeword**.

Keywords:- *Huffman code, compression, digraph, binary, codeword*

I. Introduction

Deleting old files is not necessarily needed when one's storage reaches its limit: Burning-off one's energy (bandwidth) in an attempt to transmit a large volume of data file over a cable, wireless or

any portable device can be minimized through data compression. [1].

Data compression is a technique for reducing the number of bits required to represent a text, image, sound, video etc. for space conservation or to reduce the transmission time. [2] [3] [4]. Compression techniques are generally categorized into **lossy** and **lossless**. A lossy compression technique removes unwanted information from the data compressed, the result during encoding and decoding are not always the same and is mostly used for video, sound and multimedia images. **Lossless** compression technique ensure that no data is lost at all, during encoding and decoding. The result is always the same and the technique is mostly used for text, medical images [5], [4]. Lossless techniques will reconstruct the original file exactly from the compressed file. Various compression algorithms have been developed over the years, some of these include , run-length algorithm, Huffman code, Lempel viz code, Shannon-Fano code and Morse code; Others include arithmetic coding, Joint Photographic Experts Group (JPEG) algorithm, MPEG (Moving Picture Experts Group) algorithm, Discrete cosine transform (DCT) and Code Presentation Technique, [2], [5] Different research papers that focus on natural language text compression have been published in the past few decades: Most of these research papers focused on popular natural languages like; English, French, German, Japanese, Chinese and Arabic. [6] Not much attention has been focused on the compression of Yoruba text.

According to [3], lossless technique can be classified into **statistic** compression techniques and **dictionary** compression techniques. They all try to utilize the occurrence of the same character/string in the data to achieve compression. Statistical coding technique operates by encoding symbols one at a time, the symbols are encoded into variable length output code based on the frequency of the symbols such as Shannon- Fano Coding, Huffman coding, Run Length Encoding, Arithmetic coding etc. Dictionary coding techniques operate by replacing groups of symbols in the input text with fixed length codes e.g LZW, LZ77, Code presentation technique [2], etc. The basic idea is to replace those repetitions by references to a "dictionary" containing the original.

Figure 2 diagrammatically showing types and categories of data compression

TONES	High	Mid	Low
ACENT MARK	Bá	Ba	Bà
MEANING	Meet	Weave	Impinge upon

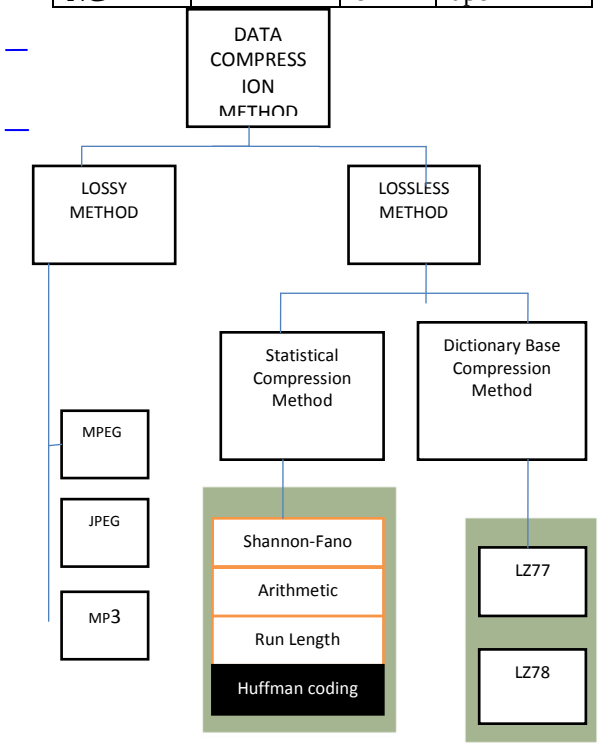


Figure1: Categories of Data Compression

II. Structure and Categories of Yoruba Words.

Yoruba word is a series of alphabet and diacritical marks. Twenty five (25) characters are used in modern Yoruba words i.e A B D E Ẹ F G GB H I J K L M N O Ọ P R S Ṣ T U W Y. Seven (7) of them are vowel A E Ẹ I O Ọ U; the remaining eighteen (18) letters are consonant B D F G GB H J K L M N P R S Ṣ T W Y. [7] It is the vowel that carries the accent mark in Yoruba text (i.e **àgbàlāgbà**), unlike in Arabic text. The diacritical marks are placed on the consonant.(i.e. يَفَسْرربِّلُجَصَّ) [8] Arabic alphabets are ظط ض ص ش س ز ر ذ د خ ح ج ث ت ب ا ع ى ء ا و ه ن م ل ك ق ف غ [9] [10]

A. Tonal structure:

The uses of diacritical marks make the tone in Yoruba language possible, it is placed directly above or below the vowels Example: **lògòlògòbángōṣè**, **àgbàlāgbà**
Syllable structure: there are three syllable structures in Yoruba language which are Consonant and Vowel (CV), Syllabic nasal and Vowel alone (V).

B. Consonant and Vowel (CV)

This is a combination of consonant and vowel to form a syllable, it's a shape of a simple verb in the language, thus: **ba, bi, ki, fo,** (meaning meet, burn, greet and jump respectively) [7]

C. Syllabic nasal

It is represented by 'n' and 'm' which also function as distinct syllables, thus: n lọ, m ba (meaning "is going", "would have" respectively).

D. Vowel alone (V)

This include regular vowel (a, e,ẹ, i, o, ọ u) and nasalized vowel (an, in, on, un etc.) each of these vowel can also function as a distinct (standalone) word.

The central figure in Yoruba language is the tone therein, three tones are basically indicated in Yoruba orthography **High**, **Mid** and **Low** [11] [7]

Table 1. Meaning of each accent mark

III. Traditional Huffman Algorithm

This section presents the traditional Huffman Algorithm for text compression. The Algorithm consists of the following steps:
Step 1: Accumulates the frequency of the characters.
Step 2: Create codes by constructing a binary tree.

1. Consider all character as free nodes.
2. Arrange the frequency of the node from highs occurred to least occurred node.
3. Assign two free nodes with lowest frequency to parent nodes with weights equal to sum of their frequency (labelling the edges from each parent to its left child with digit 0 and the right child with 1).
4. Remove the two free nodes and add the newly created parent node to the list of free nodes.
5. Repeat 2, 3 and 4 above until there is one free node left. It becomes the root of tree

IV. Modified Huffman Algorithm

This section presents the modified Huffman Algorithm for Yoruba text compression. The modified Huffman algorithm is as shown below:

1. Input your characters.
2. Check for "GB" in the characters.
3. If "GB" is found then separate "GB" as a single character.
4. Count the number of appearance of each character (frequency accumulation).

5. Assign codes to Characters by constructing a binary tree.
    1 Make a list of all characters with their frequencies (free nodes).
    2 Arrange the frequency of the node from highs occurred to least occurred node.
    3 Merge two free nodes with lowest frequency to a parent node with weights equal to sum of their frequency (labelling the edges from each parent to its left child with digit 0 and the right child with 1).
    4 Remove the two free nodes and add the newly created parent node to the list of free nodes
    5 Repeat 2, 3 and 4 above until there is one free node left. It becomes the root of tree.

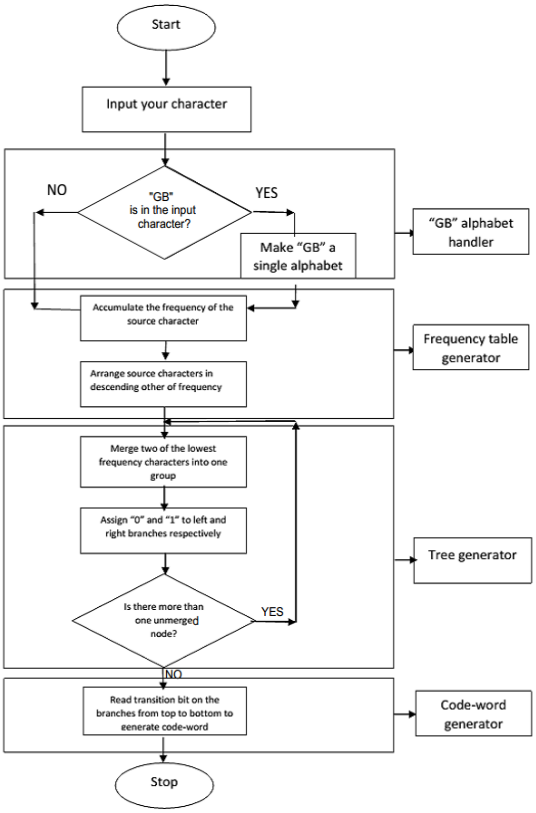


Fig 2: Improved Flowchart

V. Comparative Analysis of the Traditional Huffman Algorithm and the Modified Huffman Algorithm

The main difference between the two algorithms is that; the traditional algorithm sees the digraph "GB" as a separate alphabet but the modified algorithm recognised it as a single alphabet as it appeared in Yoruba morphology. This gives a better compression as it is seen in step2 and step3 of the improved algorithm.

Another difference is that the new algorithm caters for diacritics marks in the Yoruba text while the old algorithm does not. The steps involve in the improved algorithm is given thus:

Step1: Input your text: This is where the software will accept value from the user via keyboard or any other means

Step2: Check for "GB" in the characters: The alphabet (diagraph) "GB" is checked for from the text inputted and if found go to step3 if not go to step4.

Step3: This step will ensure that all digraphs **"GB"** in the input text is taken for a single alphabet. it will be handled as a single entity only If G is the alphabet that directly precede B, then, it is seen as "GB" but not as a "G" and "B" as a separate alphabet. "GB" will be assigned a unique code which will be different from the code which will be assigned to "G" and "B" if found in the inputted text, "GB" as a single character.

Step4: Here we count the numbers of appearance of each character: This place work internally (data abstraction), it will count all the occurrence of each character and store them into an array of character.

Step5: Create codes by constructing a binary tree: This is where the character array will be constructed and grouped as a binary tree node. The steps below support the process of building the tree. It allocates binary bits to each symbol. If the three grows to left side it is assigned '0' and if it grows right it is assigned '1'. This is where the repetition of the process that will give result takes place.

1 Consider all character as free nodes: This will first of all classify all node as a free node to make the binary tree accurate.
2 Arrange the frequency of the node from highs occurred to least occurred node: This is where the software will sort the entire character array in descending order (Highest to Lowest).
3 Assign two free nodes with lowest frequency to a parent node with weights equal to sum of their frequency (labelling the edges from each parent to its left child with digit 0 and the right child with 1).
4 Remove the two free nodes and add the newly created parent node to the list of free nodes.
5 Repeat 2, 3 and 4 above until there is one free node left. It becomes the root of tree.

VI. APPLICATION

It was discovered from the literature review that Huffman algorithm is very rich in text compression, particularly natural language. This section presents a modified Huffman algorithm as a methodology to design data compression model for memory optimization and to minimize the size of Yoruba text during transmission. This modified Huffman algorithm work in such a way that it will assign a binary code to Yoruba symbols as it is present in valid Yoruba

morphology (alphabet,(vowels and consonants), diacritical marks which indicate tones, digraph and space, are inclusive). The digraph "**GB"** is seen as a single alphabet which was neglected in the Traditional Huffman algorithm. A function is created in the algorithm which handled all digraphs in the text. This makes Yoruba words have a better compression than any previous algorithm.

The modified Huffman algorithm is efficient for Yoruba text compression, not just in terms of compression rate and compression ratio, but also save the context of Yoruba text been compressed. The following are some measurements used to evaluate the performance of the modified algorithm.

a) Space Saving

Spacing space is defined as the reduction in size relative to the uncompressed size the result is usually given in percentage. Mathematically; [12]

Space Saving = (1 -) %

b) Compression Ratio

Compression ratio (CR) is the ratio between size of compressed file and the size of source file [12]

CR= $\frac{Size\ after\ Com}{Size\ before\ Com}$%

c) Compression Time

During coding process, the program displays the Compression Time ( ) in terms of speed as at when coding begins and ends. Although, it's a function of computer performance, and it is measured in milliseconds. [13]

==

VII.

oding Process

The coding process assigns varying length code-word to symbols in the message. Unlike ASCII and UNICODE which are known as a fixed – length code, they use 8-bit code to represent each symbol in the text. if the phrase **"àgbàlāgbà"** is to be compressed using standard 8-bit ASCII code, it will thus take 8*7= 56 bits. But the use of Huffman algorithm can help compress the message in just 10 bits. We know that every character is stored in the computer in binary format; each character has a code called ASCII Code which is a sequence of **0s** and **1s.** The ASCII code of each character is represented in 8-bit which is relatively not suitable to memory optimization. It will be of a great benefit to use variable-length code for memory optimization, therefore, the method of Huffman coding is adopted.

These codes generated from the Huffman tree in such a way that every internal node (parent) I of the tree T has left L and right R children, and are assigned **0** and **1** respectively.

1. tart at the root, assign **0** to left child and **1** to the right child
2. R epeat the process down the tree
3. T o generate the code for the corresponding symbol in the text, walk through (traverse) the branches of the tree from the root to the leaf. 0 is outputted anytime there is a move to left and 1 is outputted anytime there is a move to right.

The diagram shown below gives the binary tree structure of the example of the word "**àgbàlāgbà**" by merging the frequency of two least occurred symbols in the leaves node together, resulting into a new node (internal node) whose children are the 2 nodes that their frequency was merged together, such that, new node is equal to the sum of the two node's frequency merged together. Until the root of the tree is formed.

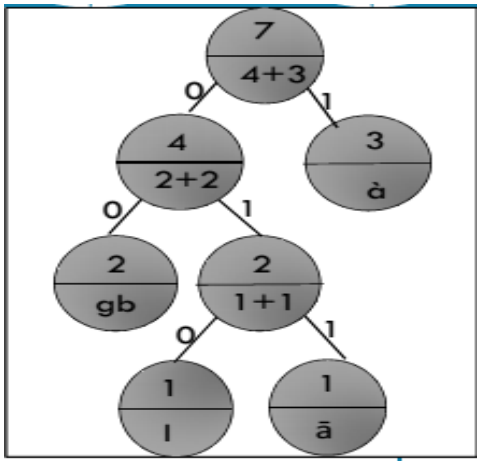


Figure 3: Example of a tree structure.

C

This will generate a short code-word to symbol that are more frequent and longer code-word to symbol that are less frequent. **1001011010001**

Here, the text is coded with just 13 bit compare with 56 bit of ASCII.

Note: that Yoruba is a tonal language and is a function of the diacritical marks placed under or above Yoruba alphabet. Same alphabet with different diacritical mark will be represented with different code-word. i.e **à** and **ā** are not the same in tone therefore they have different code-word.

VIII. Experiments and Result

Two groups of experiment were carried out on Yoruba text. The first group of experiments were aimed at comparing the efficiency of variable length Huffman techniques over fixed-length (ASCII) code for the compression of Yoruba text. The second group of experiments were aimed at comparing and evaluating the efficiency of the modified Huffman algorithm over the traditional algorithm on Yoruba text. S Several files of different sizes were tested. These files were taken from different sources on the internet.

A. Experiment One

The first experiment is performed to check the efficiency of Variable-length Huffman technique (proposed) over fixed-length (ASCII) in Yoruba text compression. The technique was applied on all the categories of Yoruba text (text with diacritical marks and text without diacritical marks). The total number of symbols present in the message is evaluated and how they will be represented in the memory using fixed and variable length code.

Table 2 shown below gives the analysis of the result of the compression bits using variable-length and fixed-length techniques for the two categories of Yoruba text.

Number of Symbols in The Text	Text Categories	Fixed - Length (ASCII)	Variable-Length (Improved Algorithm)	Compression Ratio (New Algorithm)
43	with diacritical marks	344 bits	194 bits	43%
37	with diacritical marks	296 bits	162 bits	45%
37	without diacritical marks	296 bits	148 bits	50%
585	with diacritical marks	4680 bits	2906 bits	38%
585	without diacritical marks	4600 bits	2567 bit	44%
545	English text	4680 bits	2577 bits	45%

Table 2: Result of Experiment to check the efficiency of the Modified Algorithm.

B. Experiment Two

This second group of experiment compared and evaluate the efficiency of the new algorithm over the old algorithm on Yoruba text. Various Yoruba text file of different size and categories are extracted from online document. The files are tested with the new algorithm and the result is evaluated and compared with the old algorithm using the same file size.

Text Size	Text Category	ASCII	Traditional Huffman algorithm	Modified Huffman algorithm	Compression ratio(Traditional Huffman algorithm)	Compression ratio (Modified Huffman algorithm)
1186bits	With diacritical marks	6512	4877 bits	720 bits	37%	38%
1166bits	Without diacritical mark	5440	3878bits	641.75bits	41.2%	44%
30bits	With diacritical mark	184	102bits	11bits	57%	63%
	Without diacritical mark	120	56bits	10.2bits	65%	65%
78 bits	With diacritical mark	408	268bits	40bits	46%	48%
	Without diacritical mark	312	186bits	37bits	40%	52%
1168bits	English text	4824	3220bits	643bits	44%	44%
86bits	English text	344	237	48.5	43%	43%

Table 3: Result of Experiment to check the compression ratio of the improved algorithm.

Looking at the compressed files, one will discover that the improved algorithm has a better compression ratio on Yoruba text file and it has the same compression ratio compared with English text file from table 4.2. When a Yoruba text file of 1186 bits with diacritical marks was compressed with the traditional and the improved algorithm, 37% and 38% compression ratio was achieved respectively. Also a Yoruba text file of 1166bits without diacritical marks was compressed with traditional and the improved algorithm and the compression ratio is 41.2% and 44% respective. A file of 30 bits with diacritics was tested using the traditional and the improved algorithm, compression ratio of 57% and 63% were obtained respectively. The same 30 bits file without diacritics were tested and the result for traditional and improved was the same, (65%). Another Yoruba text file of 78 bits with diacritics were tested using the traditional and the improved algorithm, the result were 46% and 48% respectively. 78 bits file without diacritics, the result obtained for traditional and the improved were 40% and 52% respectively. Looking at the English text file of 86bits using old and new algorithm has the same compression ratio of 43%. Another English text file of size 1168bits is compressed using the new algorithm and the result is compared with the old algorithm, they both have compression ratio of 44%. The implication is that the same result may be obtained if one is compressing a text without diacritical marks and without the digraph “GB”.

IX. Conclusion

Data compression is expedient for space conservation and bandwidth management in this information age. Yoruba text in particular is targeted in this research. This research work described how Yoruba text is compressed using an improved Huffman algorithm as one of the standard compression techniques. It shows that Yoruba texts with diacritical marks are better compressed with

the improved algorithm than the old algorithm that was described in the literature review.

This dissertation has been able to achieve the goal set for it by designing an efficient means of minimizing space and improving the Huffman algorithms.

The improved Huffman algorithm gives a better compression ratio when used on Yoruba text that contain diacritic marks (tones) and diagraph “GB”, It has the same compression ratio when used with English text and Yoruba text without diacritics and the alphabet “GB.”

Some of the traditional Huffman algorithm do not have consideration for diacritical marks if found in the Yoruba text and the output is not the same with the input text which should be lossless. The improved algorithm recognizes the tones in Yoruba text and is indicated by diacritical marks. The improved algorithm also recognizes “GB” as a single character which was seen as a separate character in the old algorithm. It save memory space by adopting variable length code instead of fixed length code. Yoruba text file where the digraph “GB” appeared more often tends to have a greater compression ratio.

Some Yoruba files on the digital media consist of mixed contents (text, audio, video,) a single technique may not be suitable for it compression if its applied on mixed Yoruba content, therefore, for further research in the future, it will be of great advantage if this algorithm is combined with other compression algorithm so that a better compression ratio is achieved regardless of the Yoruba file type.

VI. REFERENCES

[1] D. S. Agus, "A New Algorithm for Data Compression Optimization," *International Journal of Advanced Computer Science and Applications (IJACSA)*, pp. vol 3, NO 8, Pg 14-17, 2012.

[2] D. Oluwade, "A binary text compression algorithm based on partitioning,," *Advances in Computer Science and Engineering*, vol. vol.3 No.2, pp. pp. 165-174, 2009.

[3] S. Shanmugasundaram and R. Lourdusamy, "A Comparative Study of Text Compression Algorithms.," *International Journal of Wisdom Based Computing.*, Vols. Vol. 1, No. 3,, pp. Pp. 68-76, 2011., 2011.

[4] S. R. Kodituwakku and U. S. Amarasinghe, "Comparison of Lossless Data Compression Algorithms for Text Data.," *Indian Journal of Computer Science and Engineering*, Vols. 1 No 4 416-425., pp. Vol.(1) No.(4), pp (416-425), 2012.

[5] P. Shrusti and Y. Chaudhary, Data Compression Methodologies for Lossless Data and Comparison between Algorithms., International Journal of Engineering Science and Innovative Technology (IJESIT), 2013, pp. 2, Issue 2,.

[6] Moronfolu, Adeniyi; Oluwade, Dele, "An Enhanced Lzw Text Compression Algorithm.," *Afr. J. Comp. & ICT*, Vols. vol. 2, No. 2,, pp. Pg 13-20, 2009.

[7] A. O. Ogundepo, Contrastive Study Of English And Yoruba Morphological Systems-Implications For Nigerian Teachers And Learners Of English.International Journal of English Language and Linguistics Research., Vols. Vol.3, No.4, , 2015..

[8] E. A. Jrai, "Efficiency Lossless Data Technique For Arabic Text Compression," *International Journal of Computer Science & Information Technology (IJCSIT)*, Vols. Vol 6, No 5,, pp. 65-71, 2014.

[9] A. O. Ibraheem, "A Concept Paper on a New Model for Automatic Diacritic Restoration," 2014. [Online]. [Accessed 5 7 2016].

[10] O. C. Abikoye, A. G. Akintola and S. Ibrahim, "Design and Implementation of a Yoruba Language Crytosystem.," *Ilorin Journal of Science.*, Vols. Vol 1, No. 1, p. 50 – 60, 2014..

[11] A. Bamgbose, Yoruba Orthography booklet: outstanding problems of Yoruba orthography., Ibadan: Ibadan University Press, 1976,, pp. 15-27..

[12] P. Shrusti, Y. Chaudhary and J. Joshi, "Data Compression Methodologies for Lossless Data and Comparison between Algorithms," *International Journal of Engineering Science and Innovative Technology (IJESIT)*, vol. Vol.2 , no. Issue.2, p. pp.146, 2013.

[13] C. David, "Data Compression/Evaluating Compression Effectiveness," 2016. [Online]. Available: https://en.m.wikibook.org/wiki/Data_Compression/Evaluating Compretion Effectivess.

[14] K. Martha, Understanding English Grammar, New york, USA.: Macmillan Publishing Company., 1994.

[15] O. O. Adebusola, Mo on ko, mo on ka, mo on so. know how to write it, read it, speak it, Bis Bus International, 2010.

[16] G. E. Blelloch, Introduction to Data Compression,, Computer Science Department Carnegie Mellon Universit., 2010, pp. 22-41.

[17] Adelabu, S.B, "A Contrastive Analysis of Adjectives in English and Yoruba," *In International Journal of Education and Research*, vol. 2, p. 4, 2014.

[18] I. A. Ogunbiyi, The Search For A Yoruba Orthography Since The 1840s:, https://www.jstor.org/stable/25653397, 2002, pp. 87-102.

THE IMPACT OF ANDROID MALWARE ON MOBILE HEALTH APPLICATIONS (mHealth Apps) SERVICES

Benjamin Aruwa Gyunka
Directorate of Information and Communication Technology
National Open University of Nigeria (NOUN)
Abuja, Nigeria
gyunkson@gmail.com

Abikoye Oluwakemi Christiana
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
Kemi_adeoye@yahoo.com

***Abstract*— Sophisticated and advanced malware are being created on a daily basis and have become one most stealthy and dangerous attack approach used against critical information technology infrastructures in recent times. This paper seeks to expose Android malware attack strategies and concealment techniques and the consequent impacts on the security and privacy of electronic healthcare (eHealth) systems, particularly mobile health applications (mHealth apps). The main aim is to reveal the danger ignorant and vulnerable users, especially the sick and desperate, are exposed to when it comes to issues of Trust and Privacy arising from the use of these prevalent technology. A detailed analysis, based on secondary literatures, of various malware propagation methods and infection vectors was conducted. The research revealed that the healthcare industry information technology infrastructures experience the most of cyber attacks in recent years above other industries. With the adoption of ICT in Medicare, hospitals now accrues and stores more sensitive and personal data about their patients, more than accrued by banks and employers, but gives little or no attention to the security of their information systems. Malware attacks poses serious threats to information management and can distort critical health services such as the control of communicable and non-communicable diseases. The paper concludes by recommending for extensive peer-review of mHealth apps and for intensive measures to be taken by hospitals in implementing and continuous monitoring of a secured ICT infrastructure.**

Keywords*— *Android, Malware, mHealth, eHealth, Information Security, Communicable Disease, Non-Communicable Disease, information infrastructure, Operating System (OS)

I. INTRODUCTION

The advent and rapid growth and adoption of information and communication technology (ICT) have swept over almost every sectors of human endeavour. Android is the most popular information technology platform in circulation today among both mobile devices and desktop computers [1]. Mobile devices, particularly the Smartphones and Tablets, have grown to become the closest companion and the first point of contact to their users in almost every situation. The Android operating system is more prevalent with a proliferation rate of 86.2% above all others mobile platforms [2] and [3]. Table 1 shows the different proliferation rates of various mobile platforms.

Table 1: Mobile platform proliferation rate [3]

Operating System	2Q16 Units	2Q16 Market Share (%)	2Q15 Units	2Q15 Market Share (%)
Android	296, 912.8	86.2	271, 647.0	82.2
iOS	44, 395.0	12.9	48, 085.5	14.6
Windows	1, 971.0	0.6	8, 198.2	2.5
Blackberry	400.4	0.1	1, 153.2	0.3
Others	680.6	0.2	1, 229.0	0.4
Total	344, 359.7	100.0	330, 312.9	100.0

However, the emergence of information technology, and the popularity of the Android platform has also come along with targeted attacks against both the information stored in these technologies and the technological systems themselves. Confidentiality, Integrity, Availability, and

Non-Repudiation is the core Model designed for protecting and securing information systems. This model has however been under persistent and continuous attacks by various cyber criminals who are interested in either stealing or sabotaging vital information or information assets for different reasons. Sophisticated and advanced malware are being created on a daily basis and have become one most stealthy and dangerous attack approach used against critical information technology infrastructures in recent times.

Information and Communication Technology is used as a great tool that increases productivity, reduces cost, and safes time across Institutions, especially the health sector [2]. Android devices are very strong and viable tool for accessing and communicating sensitive health related issues for their users [3]. Android mHealth apps are intended to serve as on-demand assistance. [2] outlined a list of top ten Android mHealth applications to include:

- CareZone: - an all-in-one tool that enables users oversee their family's medication and additional required doctor's instructions
- Doctor on Demand: - tool that enables users connect and board certified doctors in their area
- Doximity: - platform for inter doctor relations that enable them share and communicate about patient care and relevant medical news
- GoodRx Drug Prices and Coupons: - enable users compare prices and find better deals for their medications
- mySugr Diabetes Logbook: - tool that enable diabetics keep track of their health conditions
- Read by QxMD: - app that deliver medical related news, the latest medical research, topics reviews, and other medical-related literature
- Epocrates: - It contains lots of info about prescription and over-the-counter drugs, including proper dosage, the effects when combined with other drugs, possible side-effects and contraindications (factors that increases risks)
- WebMD: - it is a consult on the go app for medical questions that need answers but can't get to the doctor
- Medscape: - designed to provide vital medical information

Considering the mission of the healthcare industry to save lives, one would think that they will be free from cyber attacks. It is very easy to conclude that cyber criminals will not be interested on attacking health digital infrastructures and the vulnerable and desperate patients needing medical assistance. However, it is in the nature of criminals to take advantage of any kind of existing vulnerability and weakness to carry out their exploit against any individual or organizations. Cyber criminals are more subtle and swift at exploiting any given vulnerability found on an internet network or any information systems. Malware creation is one of the most subtle and technical attack approaches adopted against the Android platform. Given that mobile devices are mostly personalised, very wealthy and influential patients, doctors and other professional health providers are the prime targets of this malware attacks. This paper seeks to take a deeper look into Android malware and their effects on providing efficient digital healthcare services especially in the area of managing information and controlling communicable and non-communicable diseases.

II. RELATED WORK

Smartphones and Tablets are cutting-age technological advancement of the 21st Century which has unequivocally introduced rising security challenges in the field of computing. The emergence and proliferation of these Hand-held Devices is almost reaching an epidemic proportion [6]. The world's population study reveals the existence of about seven billion people in the Universal [7] and more than 87% of this population already have been penetrated by the mobile market [8]. These devices have the in-built capabilities of remaining constantly connected to the internet (due to the possession of features such as Wi-Fi, voice, data, GPS, etc.) and have continued, at an exponential rate, to win the confidence and trust of their users. Their emergence has simplified lots of social networking activities, commercial and banking transactions so much that they have become an integral part of our lives. [9, p. 37 & 38] observed that the trust that people have in their handheld devices has by far outweigh the trust they have for their fellow humans. These devices hold more of their intimate personal secrets and corporate life that no one else has access to. This has created wider vulnerable environments which have been rapidly exploited, on a daily basis, by different category of malicious users and hacktivist. Malicious activities and attacks have continued to grow exponentially against mobile devices especially those running the Google Android OS, which is an open-source software stack, given their popularities and influence in the economy today [10] and [11]. The threats confronting these devices can be Application Based, Network Based and Web Based [12]. Attacks through malicious applications are more prominent against the Android Platform [13]. Thus, mHealth applications and eHealth have serious privacy and security issues [14], [15], [16], [5], [17], [18], and [19]. It has become very imperative therefore to critically research Android malware, their behaviours and concealment strategies which they deploy against information systems.

III. THE ANDROID FRAMEWORK

The android platform is divided into three main parts namely, the Software Development Kit (SDK); the Operating System (OS); and Applications [20]. This platform is built upon Linux 2.6 kernel and was originally under the supervision of the Open Handset Alliance [21].

Because of this open source nature of the platform, lots kinds of developers were attracted to it [22]. The Linux kernel is the bedrock of the Android platform. Every essential components and services, such as process management, security, memory management, and lots more are only enabled and controlled by the Linux kernel. The Android framework compromises of different layers which begin with the Linux Kernel at the very bottom, forming the core of the operating system. Another key component in this framework is the Dalvik Virtual Machine. The Dalvik VM creates an enabling environment that allows processes to run separately [23] and [24], each having its unique process id either known as User ID (UID) or Group ID (GID) and also having unique storage [25]. The existence of UID and GID disallow application processes from accessing or sharing the storage or resources of each other without strict permission granted [22].

Software Development Kits (SDK) is a Java toolbox which Google have released freely to the public to enable anyone creates any Android application at will without restrictions [22]. One of the most important tool found in the SDK is the Android Debug Bridge (ADB), it helps to interface an Android device with a computer system to enhance effective communication and application development [27]. Through the adb the application developer is able to access the command shell via the computer system and other activities such as application installation/uninstallation, reading of logged files, and file transfers and be easily made without obstructions [20].

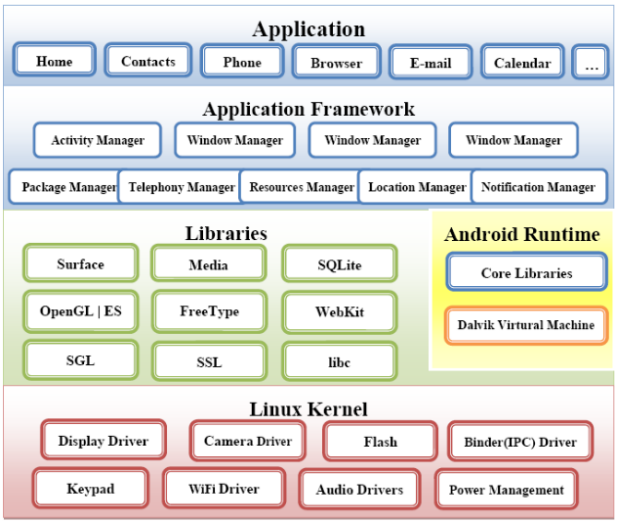


Figure 1: Android Framework [26]

The operating system (OS) makes use of the concept of sandbox to manage applications and processes execution. It uses also the SQLite database including several file system formats like YAFFS2 (Yet Another Flash File System 2), RFS, and EXT4 [20] and [25]. Android Operating System has now extended beyond just running on Smartphones and Tablet but it is found on net books, e-readers, cars, and home and hospital appliances [22].

A. Android Data Storage Directories

Mobile phones are traditionally known to store important data in SIM-cards or device's internal memory [28] or even with the mobile network providers [29] and in an external memory card. Android's internal memory is called NAND memory. The Android stores data in directories as shown below [30];

- Share Preferences: - the data here is stored as key-value pairs which found in the directory "/applications/data/shared_pref"
- Internal Storage: - this storage is also known as the NAND flash, it contains mostly private sensitive data
- External Storage: - the directory for this storage is '/sdcard', it contains all public information and does not necessarily have strong security measures
- SQLite: - SQLite is a structured Android database with the directory '/data/data/Package/database

The concept of rooting an Android device is a delicate and very risky process. Rooted devices are highly vulnerable and exposed; malware can easily penetrate the filesystem, defiling all security measures, to access all these data. More so, the adb tool can be used against the rooted device by an attacker to extract almost all allocated data on the system [31].

B. Android Application Framework

All Android applications are created and compiled as Android Package (APK) file [32] and [33]. The APK file is simply a ZIP format archive file that is renamed to have an .apk extension [34]. APK files would normally have to be unzipped, as the case with normal zip files, in order for its contents to be extracted. The content consists of a Dalvik executables, resources, native libraries and a manifest file, and is usually signed by the developer of the application using self-signed certificate [33] and [32]. In particular, the APK file would usually contain two folders (META-INF and res) and three files (classes.dex, AndroidManifest.xml, resources.arsc) as can be seen in figure 2 [32]. The classes.dex and AndroidManifest.xml are the most delicate and important components of the APK file which are usually the high targets of malware creators [35] and [36]. The classes.dex is the Dalvik Virtual Machine executable file which contains the main working code of the application. That is, the payloads of an app is created and defined in classes.dex file. The AndroidManifest.xml provides semantic-rich information about the application which includes the version and required permissions governing an app's operations. AndroidManifest.xml also contains four basic components which are Activities, Services and Broadcast Receivers and Content Providers. It

is important to note that without the information in manifest file, it is impossible for an application to run on the Android OS [33] and [35]. Attackers target these two files to either inject malicious code [36] or alter the permissions for nefarious purposes. Activity is simply an element of the user interface which represents a screen. Service works at the background and can run indefinitely.

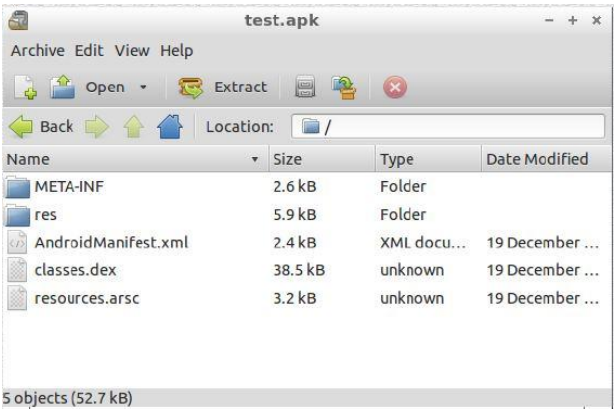


Figure 2: An APK unzipped file displaying its contents [34]

Broadcast receiver is an application component which receives broadcast messages and starts a workflow according to the obtained message. Content Provider is the component that provides an application with the abilities to store and retrieve data. It also enables sharing a set of data with another application. Importantly, the different components contained in an Android Package makes an app not to have a central entry point like in the case of Java programs that have main method. Not having a central entry point implies that all components (except for broadcast receivers which may be dynamically defined also) must necessarily be declared by the application's author in the AndroidManifest.xml file [33].

IV. ANDROID MALWARE

Malware is any software used to disrupt computer operation, perform unauthorized functions, gather sensitive information or gain access to private information systems [37]. Malware can range from a simple nuisance such as pop-up advertising to an intrusion causing severe damage such as stealing passwords and data or infecting other machines on a network. Malware appear in different forms such as Viruses, Spyware, Adware, Nagware, Backdoors, Rootkits, Trojans Horse, Botnets keyloggers, spammer and Flooder, Zombie, Auto-rooter, Logic bomb, and Worms [38] and [39]. Within the mobile platform, malware are able to propagate between devices using infection vectors such as Bluetooth, MMS, SMS, Internet networks, executable files, emails and web vulnerabilities [1]. Other infections vectors include social engineering, synchronization/USB, WiFi, Device Rooting, Near Field Communication (NFC) Peripheral memory cards, Drive-by downloads, and Google Play Store [38].

Malware can be classified according to their behaviour and payloads [32]. They can also be classified in accordance to the way they spread or propagate [38]. There is the group that need to be triggered by a program or by some users' activities to enable them replicate and spread (Such as viruses), the second group of malware (Worms) are self replication ability and can automatically scan for vulnerable victims through internet networks or any medium in contact with them. The third category evolves into exploits and unknown vulnerabilities. They employ multiple attack vectors in order to replicate. These last categories are known to strike faster, preventing timely intervention by security administrators. Some notable Android propagation technique include Drive-by Download, Repackaging [40]. Devices that are infected with Malware, usually, are largely under the control of the attackers. They can provide an unauthorized gateway for the attacker to be able to access and steal very personal sensitive information such as Bank details, confidential messages, files, photos and perform online/click fraud and phishing. The attacker will also be able to remotely control the device by placing illegal phone calls, sending spam emails, SMS, and MMS. "Fake Player" was the first Android Malware discovered in 2010 then subsequently others such as Pjapss, Hippo, Geinimi, DroidDream, Droidkungfu and Basebridge and NotCompatible, which is an Android Web Proxy Bot. KorBanker is another example of a nasty Android malware found in South Korea which targeted several popular South Korean banking apps. The most recent Android malware released in the wild is Stagefright [41]. Android malware are always on the increase and are becoming tougher and resistance, through different concealment techniques and sophistications, to existing detection methods and models [38]. Android malware behave differently in accordance to their stated payloads and requires different requisite techniques to put them in check [42].

The kind of attacks deployed by malware includes unauthorized access, masquerading, denial of service, annoyance attack, and eavesdropping [38]. Malware have evolved with advanced attack techniques against the Confidentiality, Integrity, and Availability Model of information security. The confidentiality, integrity, and availability of information and information assets can no longer be guaranteed with the sophistication evolvement in malware developments and attacks. The main idea behind malware development is to generate revenue for the developer (attacker) while generating cost for the user (victim). This is achieved as the malware are targeted at stealing financial information (Credit/Debit Card details, Mobile banking Login credentials, and Hijacking e-commerce transaction sessions) and other very personal information via click fraud, extortion and espionage. Usually, malware are mostly operating systems and hardware specific, but recent development has seen the creation of advanced malware that have the capacity to

operate across platforms –especially between Android and Windows which are the most popular mobile and desktop platforms [38]. This simply implies that there exist malware that can now successfully operate across two or more distinct operating systems.

A. Malware ConcealmentStrategies

Malware advanced sophistications are aimed at enabling escape and withstand any existing detection methodologies. The ability of a malware to hide its presence and escape detection is also termed as concealment strategies. The goal of this strategy is to enable the mal-code survive longer without detection in order to have ample time available to replicate and infect more devices. That is, the technique increases the life span of a malware between infection and detection phases. Also, these techniques aim to make the analysis part of malware more difficult for the antimalware developers. [38] noted that malware concealment techniques can be classified into two main groups. The first group bypasses anti-virus or anti-malware signature detection by implementing a set of methods to enable it hide itself. Group two are not only able to hide their presence from detection but also can make active effort to hinder the detection and analysis of the code using anti-emulation, anti-debugging techniques and anti-disassembling techniques and also evade signature-based detection methods. These second groups of malware are known to implement aggressive coding techniques for self-protection while the first group does not have the ability to attempt to defend itself. The first class of malware adopts a passive concealment approach while the second class deploys an active approach. Thus, malware passive concealment techniques involve code obfuscation, entry point obfuscation, encryption, Packing, Oligomorphism, Polymorphism, and Metamorphism. Active strategies for malware concealment involve Anti-emulation, Anti-debugging, and Anti-disassembling.

1) Active Approaches

a) Anti-emulation: Across all platforms, the use of virtual machines and emulation are very vital aspect of malware analysis because they provide a safe and accurate environment for the analysis and evaluation of real life behaviour of malware and its impact on the targeted system. The use of emulation as safe haven for malware analysis by security analyst has also received counter response from malware creators who have in turn implemented anti-emulation techniques to counter analysis and evade detection.

b) Anti-debugging: An anti-debugging technique is mostly deployed to trick debuggers, via code obfuscation technique, which then makes the job of anti-malware analysis much more difficult.

c) Anti-disassembling: Anti-disassembly technique employs special code within a file that trick the disassembly and analysis tools into producing incorrect programme listings. This technique provides an additional layer to the extensive array of armour available for malware creators. When implementing anti-disassembling in their code, malware authors have two common goals they strive to achieve – that the function should not be easily automated and that the source code should not be available until the code runs.

2) Passive Approaches

a) Obfuscation: Obfuscation is a method adopted by code developers to enforce a security policy targeting at protecting the code from being analysed and understood. Code obfuscation is a deliberate modification of the code with the intention of increasing it detection difficulty by anti-malware and understanding by humans. Code obfuscation can successfully be done through dead code insertion, code transposition and registers reassignment (this technique is extensively used by metamorphic viruses). Entry point obfuscation is another technique which viruses hided inside executable files adopt to call their malicious code. When a file is infected by a virus, the virus seeks for a way to take control of the file and inject its malicious code. the payload of such malware are usually hidden inside unusual places of the executable file and thus tries to distract the anti-virus or anti-malware engine from investigating the file. Entry point obfuscation is considered as an anti-heuristic method and is often deemed hard to detect, disinfect and remove.

b) Encryption: This is a technique adopted by most malware that involves scrambling their contents or code. The malware will usually creats a random encryption key which it then stores in the body of the malware and then encrypts its remainding part. It comes alive when invoked and it then decrypt itself for action [39].

c) Packing: author and affiliation lines of affiliation 1 and copy this selection.

d) Oligomorphism: Insert one hard return immediately after the last character of the last affiliation line. Then paste down the copy of affiliation 1. Repeat as necessary for each additional affiliation.

e) Polymorphism: Place your cursor to the right of the last character of the last affiliation line of an even numbered affiliation (e.g., if there are five affiliations, place your cursor at end of fourth affiliation). Drag the cursor up to highlight all of the above author and affiliation lines. Go to Column icon and select “2 Columns”. If you have an odd number of affiliations, the final affiliation will be centered on the page; all previous will be in two columns.

f) Metamorphism: Metamorphic malware employs the technique of swapping registers in a system or CPU from one form or generation to another in order to prevent conventional and known signature and pattern matching from working. This malware have the ability to re-programme itself as it evolved after each successful

infection. It does this by completely rewriting itself at each iteration which increases the difficulty of detection [39]. Metamorphic malware is divided into two main groups. The first group stages out the communication capability of the malware whether it is capable of open-world communication (downloading further plugins and updating itself with new features) or closed-world communication with no communication capability. The second group is characterized by the transformation capability of the malware into different behaviour and appearance. Metamorphic malware is believed to be more complex than any other malware [38], for the posses a strong ability to dynamically disassemble themselves, transform their code, and reassemble themselves into an executable form.

B. Trust and Privacy in e-HealthCare

Health problems drive people to seeking for varied information from and through various sources in order to deal with their health challenges [14]. With the advent of the internet (Information Technology) and particularly mobile devices, the amount of information within the finger tip of health patients and professionals is of great volume so much that they are faced with the challenge of Trust and Security for these technologies [14] and [15]. The wide spread and adoption of Smartphones and the corresponding large volume of various mobile health application in circulation [44], more and more people now share their sensitive and personal healthcare information using these apps [45]. Users' private information is left at the mercies of the developers of these apps as the common users usually do not know how these applications manage or use their data. This poses serious security, privacy and trust issues to the general public. This is because there is a lack of substantive reference information for consumers to base well-informed decisions about whether or not to adopt the applications they review and to ascertain the validity of the information provided by these e-health solutions [14]. With the exponential rise of malicious mobile apps, it is becoming increasingly difficult to guarantee the quality of these m-health apps, particularly in respect to serious concerns regarding the safety and validity of the information the common users' access through them [46] and [47]. [15] observed that up to date, there is just little scientific proof that mobile health applications (m-Health apps) can be trusted and used safely.

A most recent report shows that among all America's critical infrastructures, the healthcare industry is the most targeted and plagued by perpetual persistent attacks from numerous unknown malicious hackers whose intents is to exploit vulnerabilities in their insecure and antiquated networks in order to exfiltrate patient health records [17]. The record also reported that approximately 47% of United State populations have experienced a compromised in their personal healthcare data over the last 12 months. This goes on to show that the healthcare sector suffered from the most recent data breaches. It is amazing the broad amount of sensitive patients' data collected by hospitals, this includes address, social security numbers, spouse, children, credit cards and bank accounts [17]. Research has shown also that hospitals accrues a surprisingly wide amount of information and stores it in one (often-vulnerable) system [17]. The domain of medical mobile application development has thus become a high target for malware developers who are either after the susceptible users for personal gains or for espionage. It can clearly be stated that your hospital has a greater and broader amount of your private data than your employer or your bank does [17]. The Healthcare industry is extremely susceptible to cyber-attacks and very little fraction of them are given attention to implementing strong information security systems on their infrastructures [17].

The health sector is so engulf and very busy, channeling all their resources, to fulfilling their main vision, which is saving lives, that they give little or no attention to the security of their Information Technology Systems. Malicious attackers on the other hand have seen this selfless dedication and commitment to saving lives as a weakness. Being fully aware of this strong weakness and vulnerability in the health industry, they are making due diligent exploiting it through various attack surfaces such as advanced persistent threats (APTs), Distributed Denial of Service (DDoS) attacks, malware infections, cyber espionage, and data and intellectual property theft [16]. Thus, the Integrity, Confidentiality, and Availability of m-health users' data cannot be completely assured [18]. It therefore recommended that the health sector would need to invest in risk management based information security programs and should engage in cyber security programs that deploys a multilayered defense that protects the confidentiality, integrity and availability of information whenever it is stored, in transit, or being processed [17].

C. The Challenge of Android Security Model

The Android operating system has in-built authentication mechanisms which enables the use of passwords, tactile patterns or biometric information [20]. Most importantly, the platform takes advantage of its file system security and permission model derived from its core Linux kernel, to enforce its own internal security [33]. As can be observed from figure 1, the Linux Kernel is the lowest level of the Android framework and it is the backbone of the Android inbuilt security. Every layer of the framework has some level of security they provide. Any further security measures that is designed for this platform can only be built upon existing ones [48]. A few security challenges confront Android devices. The first is that they have limited and restricting security architecture, such as slow processing power, small storage capacity and short battery power. These limited resources make it impossible for security experts to run a complete anti-malware solution in them. The second challenge is that most existing malware analysis techniques were built to reside and run at high level

within the operating system stack [38] and [42]. The performance of these detection techniques can easily be influenced by a compromised OS in which they reside. That is, an infected device can create an unreliable environment for the analysis tools by influencing their activities [1]. Attackers are fully aware of this loophole and have successfully implemented various concealment strategies, both active and passive to invade detection. The Domain of Android malware is a virile one that requires more extensive research into novel ways of curbing the emergence of new and sophisticated malware and protecting critical information infrastructures.

V. CONCLUSION

The rapid proliferation and adoption of Smartphones, especially the Android, has seen massive development of different mobiles applications created to ease lots of human activities by reducing workloads, improve healthcare quality, improve efficiency, reducing cost and saving time [18]. Considering the important of Smartphones and mobile apps in the healthcare sectors, it can be observed that physicians and other health professionals now use them to access patient records, to view test results, and to prescribe medications. Health patients also use these mobile health apps to access and update their medical records, to monitor their health statistics, and to also view their prescription [45]. Although these technologies are very helpful and handy tools for critical information management and dissemination especially in contributing to the control of communicable and non-communicable diseases in the health sectors, strong and extensive peer-review of these apps is necessary. Stronger security and privacy measures have to also be put in place to ultimately fight against cyber attacks that will take advantage of the vulnerable health patients through malware attacks and miss-management. Priority has to be placed on Android platform because of its open nature and popularity; most of the apps developed for the platform do not require approval from anyone before they are released into the market [49]. This defines why Android malware ranges highest above malware of other platforms. If the issues of Trust and Privacy of Information and sensitive data can be addressed in the health industry, the benefits of information and communication technology will be greatly felt in regard to communicating relevant, life-saving, health information to the public.

References

[1] K. Dunham, S. Hartman, J. A. Morales, M. Quintans and T. Strazzere, Android Malware Analysis, New York: CRC Press, 2015.

[2] International Data Corporation, "Smartphone OS Market Share, 2016 Q2," 17 August 2016. [Online]. Available: http://www.idc.com/prodserv/smartphone-os-market-share.jsp. [Accessed 30 September 2016].

[3] Gartner, "Gartner Says Five of Top 10 Worldwide Mobile Phone Vendors Increased Sales in Second Quarter of 2016," 18 August 2016. [Online]. Available: http://www.gartner.com/newsroom/id/3415117. [Accessed 30 September 2016].

[4] Health, "Top medical apps for Android," 29 August 2016. [Online]. Available: http://androidapps.com/top-medical-apps-for-android/. [Accessed 30 September 2016].

[5] D. He, N. Muhammad, C. A. Gunter and N. Klara, "Security concerns in Android mHealth apps.," in In AMIA Annu Symp Proc, Washington DC, pp. 645-654, 2014.

[6] B. Mellars, "Forensic examination of mobile phones," Digital Investigation, vol. 1, no. 4, pp. 266-272, 2004.

[7] United Nations, Department of Economic and Social Affairs, Population Division, " World Population Prospects: The 2010 Revision, Highlights and Advance Tables," United nations (UN), New York, pp. 17-48, 2011.

[8] C. Knudsen, "Smartphones, tablets and the mobile revolution," 29 January 2013. [Online]. Available: http://www.mobilemarketer.com/cms/opinion/columns/14667.html. [Accessed 18 April 2013].

[9] A. Hoog, Android Forensics: Investigation, Analysis and Mobile Security for Google Android, J. McCash, Ed., Amsterdam: Syngress, 2011.

[10] B. Baskaran and R. Anca, "A Study of Android Malware Detection Techniques and Machine Learning," in MAICS 2016, pp. 15-23, 2016.

[11] S. Chatterjee, P. Kasturi, R. Reek and N. Asoke, "A Comprehensive Study on Security issues in Android Mobile Phone—Scope and Challenges," International Journal of Innovative Research in Advanced Engineering, vol. 3, no. 3, pp. 62-72, 2016.

[12] E. Charani, E. Castro-Sánchez, L. S. Moore and A. Holmes, "Do smartphone applications in healthcare require a governance and legal framework? It depends on the application!," BMC Medicine, vol. 12, no. 1, p. 1, 2014.

[13] F. Rafael, S. Julian and K. Marcel, "On the Effectiveness of Malware Protection on Android - An evaluation of Android antivirus," Fraunhofer AISEC, p. 45, 2013.

[14] W.-C. Su, "A preliminary survey of knowledge discovery on smartphone applications (apps): Principles, techniques and research directions for e-health," in ICME International Conference on Complex Medical Engineering June 26-29, Taipei, 2014.

[15] R. J. Riezebos, "Peer-reviewing of mHealth applications," University of Amsterdam, Amsterdam, pp. 6-46, 2014.

[16] B. Filkins, "Health Care Cyberthreat Report: Widespread compromises detected, compliance nightmare on horizon," SANS Institute, Bethesda, MD, pp. 2-22, February2014.

[17] ICIT, "ICIT Brief Hacking Healthcare IT in 2016: Lessons the HealthCare Industry can learn from the OPM Breach," Institute for Critical Infracture Technology, pp. 1-11, January 2016.

[18] D. Kotz, "A threat taxonomy for mHealth privacy," in In COMSNETS, Hanover, pp. 1-6, 2011.

[19] KPMG, "Health Care And Cyber Security: Increasing Threats Require Increased Capabilities," in KPMG 2015 Cyber Healthcare Survey, pp. 1-7, August 2015.

[20] M. d. L. S. Andre, C. S. Fabio, P. d. M. Laerte and T. d. S. J. Rafael, "Acquisition of digital evidence in android," in Proceedings of the 9th Australian Digital Forensics Conference, pp. 1-10, 2011.

[21] J. Lessard and G. C. Kessler, "Android Forensics:Simplifying Cell Phone Examinations," Small Scale Digital Device Forensics Journal, vol. 4, no. 1, pp. 1-12, 2010.

[22] A. Hoog, "An introduction to Android forensics," 2010. [Online]. Available: http://www.dfinews.com/articles/2010/04/introduction-android-forensics?page=0%2C0&pid=974#.UgPsHZbsZ3I. [Accessed 08 August 2013].

[23] M. Yates, "Practical Investigations of Digital Forensics Tools for Mobile Devices," in 2010 Information Security Curriculum Development Conference, pp. 156-162, 2010.

[24] M. Chow, Composer, Android Forensics Demo. [Sound Recording]. INFOSEC World Conference and Expo. 2012.

[25] C. Triplett, "Android Forensics Background, techniques and analysis tools," 2013. [Online]. Available: https://viaforensics.com/mobile-

security/htcia-android-forensics-training-presentation-march-22-2013.html. [Accessed 04 August 2013].
[26] L. Xinfang, Y. Chunghuang, C. Shihjen and W. Jainshing, "Design and Implementation of Forensic System in Android Smart Phone," in In The 5th Joint Workshop on Information Security., pp. 1-11, 2009.
[27] Developers, "Android Debug Bridge," 2013. [Online]. Available: http://developer.android.com/tools/help/adb.html. [Accessed 27 August 2013].
[28] S. Y. Willassen, "Forensic analysis of mobile phone internal memory," in Advances in Digital Forensics, pp. 191-204, 2005.
[29] D. Forte and A. Donno, "Mobile Network Investgations," in Handbook of digital forensics and investigation, C. Eoghan, Ed., Amsterdam, elsevier Inc., 2010, pp. 517-557.
[30] T. Rohit, "Android Forensics," 2013. [Online]. Available: http://resources.infosecinstitute.com/android-forensics/. [Accessed 18 August 2013].
[31] S. Kevin, "ADB Method," 2011. [Online]. Available: https://viaforensics.com/android-forensics/adb-method.html. [Accessed 18 August 2013].
[32] R. Paolo, "Developing A Next-Generation Mobile Security Solution For Android," Reykjavík University, pp. 91, 2014.
[33] Y. Zhauniarovich, Android Security (and Not) Internals, Trento: ASANI, 2014.
[34] P. O. Fora, "Beginners Guide to Reverse Engineering," in RSA Conference 2014, San Francisco, pp. 1-22, 2014.
[35] S. Rushil, "Analyzing and Dissecting Android Applications for Security defects and Vulnerabilities," Blueinfy, pp. 1-14, 2011.
[36] F. Tchakounté and P. Dayang, "System calls analysis of malwares on android," International Journal of Science and Technology, vol. 2, no. 9, pp. 669-674, 2013.
[37] M. Federico, "Malicious Android Apps - Overview, Status and Dilemmas," Andrototal, pp. 1-75, 2014.
[38] A. Nicholas, "Cross-Platform Malware Contamination," RHUL, London, pp. 1-84, 2015.
[39] R. Indrajit, "Computer Security," Colorado State University, 2013.
[40] Y. Zhou and J. Xuxian, "Dissecting android malware - Characterization and evolution," in In 2012 IEEE Symposium on Security and Privacy, pp. 95-109, 2012.
[41] N. Phil, "The 'Stagefright' exploit: What you need to know," 17 August 2015. [Online]. Available: http://www.androidcentral.com/stagefright. [Accessed 15 October 2016].
[42] R. Rahul, R. Venkiteswaran, J. B. Anoop and K. D. Soumya, "Android Malware Attacks and Countermeasures: Current and Future Directions," in Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on IEEE, pp. 137-143, 2014.
[43] A. Sharma and K. S. Sanjay, "Evolution and detection of polymorphic and metamorphic malwares: a survey," p. 1406.7061, 2014.
[44] M. Alexander, U. Philipp, W. Harald and S. Stefan, "Analyzing Privacy Risks of mHealth Applications," in Studies in health technology and informatics 221, pp. 41-45, 2015.
[45] P. Miloslava, A. Steven and C. Samir, "A Taxonomy of mHealth Apps – Security and Privacy Concerns," in System Sciences (HICSS), 2015 48th Hawaii International Conference on, pp. 1-10, 2015.
[46] B. J. Visser and W. B. Arthur, "Need to peer-review medical applications for smart phones," Journal of Telemedicine and Telecare, vol. 18, no. 2, pp. 124-124, 2012.
[47] T. L. Lewis, "A systematic self-certification model for mobile medical apps," Journal of medical Internet research, vol. 15, no. 4, p. e89, 2013.
[48] N. Elenkov, Android Security Internals: An In-Dept Guide to Android's Security Architecture, San Francisco: No Starch Press, Inc., 2015.
[49] R. Meier, Professional Android 4 application development, John Wiley & Sons, 2012.
[50] J. Morales, X. Shouhuai and S. Ravi, "Analyzing malware detection efficiency with multiple anti-malware programs," in Proceedings of 2012 ASE CyberSecurity'12, pp. 56-66, 2012.
[51] A. H. Sung, X. Jianyun, C. Patrick and M. Srinivas, "Static analyzer of vicious executables (save)," in ." In 20th Annual Computer Security Applications Conference, 2004, pp. 326-334, 2004.
[52] H. Shewale, P. Sameer, D. Vaibhav and S. Pragya, "Analysis of android vulnerabilities and modern exploitation techniques," ICTACT Journal on Communication Technology, vol. 5, no. 1, pp. 863-867, 2014.
[53] H. Gonzalez, A. A. Kadir, S. Natalia, A. J. Alzahrani and A. A. Ghorbani, "Exploring reverse engineering symptoms in Android apps.," in In Proceedings of the Eighth European Workshop on System Security, pp. 7, 2015.
[54] O. Jon and M. Charlie, "Dissecting the Android Bouncer," in SummerCon 2012, pp. 61, 2012.
[55] A. Desnos and G. Geoffroy, "Android: From reversing to decompilation," in Proc. of Black Hat Abu Dhabi, pp. 77-101, 2011.
[56] DREAM LAB Technologies, "Reversing Android Apps," DREAM LAB Technologies, pp. 61, 2011.
[57] A. D. Vicente, "Android reverse engineering: understanding third-party applications," The OWASP Foundation, Romania, pp. 1-24, 2013.
[58] N. Godfrey, Decompiling Android, Apress, 2012.
[59] S. Vanja, "Sophos Mobile Security Threat Report," SophosLabs, pp. 1-10, 2014.
[60] Y. Zhou, W. Zhi, Z. Wu and J. Xuxian, "Hey, You, Get Off of My Market: Detecting Malicious Apps in Official and Alternative Android Markets," in In NDSS, pp. 50-52, 2012.
[61] R. Adhikari, D. Richards and K. Scott, "Security and Privacy Issues Related to the Use of Mobile Health Apps," in Proceedings of the 25th Australasian Conference on Information Systems, 8th - 10th December, Auckland, pp. 1-11, 2014.
[62] F. Zubaydi, S. Ayat, A. Fadi and S. Assim, "Security of Mobile Health (mHealth) Systems," in In Bioinformatics and Bioengineering (BIBE) 2015 IEEE 15th International Conference on, pp. 1-5, 2015.
[63] C. Zhang, Z. Xihui and H.-N. Richard, "Assessment Metrics, Challenges, and Strategies for Mobile Health Apps," Issues in Information Systems, vol. 14, no. II, pp. 59-66, 2014.
[64] L. Ohno-Machado, W. Shuang, W. Xiaofeng, I. Arya and J. Xiaoqian, "Privacy, Security, and Machine Learning for Mobile Health Applications". aaas, pp. 1-13, 2014
[65] The mHealth Alliance, "Patient Privacy In A Mobile World - A Framework To Address Privacy Law Issues In Mobile Health," mHealth Alliance, pp. 116, June 2013.
[66] Y. Y. Yip, "There's an app for that: the use of mobile medical applications in clinical practice," OntariO Medical review, p. 48, 2014.
[67] "Cyber Security Management: A Review," Business Management Dynamics, vol. 5, no. 11, pp. 16-39, 2016.
[68] E. Chin, P. F. Adrienne, S. Vyas and W. David, "Measuring user confidence in smartphone security and privacy," in In Proceedings of the Eighth Symposium on Usable Privacy and Security, pp. 1-15, 2012.
[69] G. Zyba, "Mobile Malware Propagation and Defense," University of California, San Diego, pp. 115, 2013.
[70] A. John, "mobile Apps and mobile Devices Safe to Use?," NHS Ayrshire and Arran, pp. 1-23, 2015.
[71] P. CARL, "Best Android Apps For Doctors, Physicians, And Medical Practitioners," 25 May 2012. [Online]. Available: http://www.androidauthority.com/best-android-apps-doctors-physicians-health-professionals-88851/. [Accessed 30 September 2016].
[72] M. Pomilia, "A study on obfuscation techniques for Android malware," Sapienza University of Rome, pp. 81, 2016.
[73] K. W. Hamlen, "Stealthy software: Next-generation cyber-attacks and defenses," in In Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on, pp. 109-112, 2013.
[74] D. Maiorca, A. Davide, C. Igino, A. Marco and G. Giorgio, "Stealth attacks: An extended insight into the obfuscation effects on Android malware," Computers & Security, vol. 51, pp. 16-31, 2015.
[75] K. Allix, J. Quentin, T. F. Bissyandé, K. Jacques, S. Radu and L. T. Yves, "A Forensic Analysis of Android Malware--How is Malware Written and How it Could Be Detected?," in In Computer Software and Applications Conference (COMPSAC), 2014 IEEE 38th Annual, pp. 384-393, 2014.

Pattern Recognition: A Decipherable Tool for Intelligent Transportation System in Nigeria

Oluwafisayo Babatope Ayoade
Computer Science Department,
University of Ilorin, PMB 1515,
Ilorin, Kwara State, Nigeria
oluwafisayoadeb.oba@gmail.com

Bamidele A. Oluwade
Computer Science Department,
University of Ilorin, PMB 1515,
Ilorin, Kwara State, Nigeria
deleoluwade@yahoo.com

Ayorinde Oladele Idowu
Computer Science Department,
College of Education, PMB 250,
Ikere Ekiti, Ekiti State, Nigeria
aoidowu87@gmail.com

***Abstract* – this paper presents a system framework as an intelligent system for the detection, identification and recognizing vehicle plate numbers violating traffic lights on Nigerian roads using Pattern Recognition as a constructive instrument. Intelligent Transportation System (ITS) is the application of advanced sensor, computer, electronics, and communication technologies and management strategies to improve the safety and efficiency of the surface transportation system. ITS is an important implementation technique for analyzing and handling traffic control management of real time moving vehicles on roads in urban environment. The scientific discipline of machine learning or artificial intelligence with the purposes of classifying data (patterns) into a number of categories or classes is known as Pattern Recognition. The intelligent system uses the Nigerian vehicle license number plates for image processing and character recognition technology in order to identify the vehicle plate number that violates traffic lights automatically through surveillance cameras and intelligent image analytic software. The traffic control management is control by automatic signaling system where coloured lights, i.e. Red, Yellow and Green, are used for interpretation for traffic controlling operations signals. The system uses Pattern Recognition models to isolate the features of the vehicle license plate of each region to be extracted in order to correctly differentiate the identified license plate from other regions. The paper will be highlighting the various pattern recognition models that are suitable for the vehicle license plate number character recognition and pattern database matching**

Keywords— Pattern Recognition, Intelligent Transportation System, Traffic Light, Vehicle License Plate

I. INTRODUCTION

Road Transportation, in Nigeria, is one of the most common means of transportation compared to other types such as air, water and rail. With the country's technological advancement and democracy era, not minding the economy depravity, a large number of Nigerians have been afforded to own a car which has greatly increased adversely the number of road users over a decade. Technological advancement in the area of transportation has make life easy in contrast to the traditional way of transportation e.g. animals, walking, etc. which has positively increase the number of car owners. This development has also greatly increased the rate of traffic violations especially in cities and mega cities. Road accidents are supposed to be caused by various reasons ranging from drivers' carelessness, drunkenness, deplorable state of the road, etc., but it has been shown that the statistical value of road accidents has spontaneously risen over the years especially in traffic lights violations which supposed to reduce the rate of accidents. In this vein, the Nigerian government has tried in so many ways to put in place measures to control the high rate of road accidents. One of these measure is the issuance of primary identifier which is the vehicle identification number that is actually the legal license plate number unique to every car registered in Nigeria be it private, public or government registered vehicles.

Intelligent Transportation Systems (ITSs), as defined by [1], [2] are the applications of advanced sensor, computer, electronics, and communication technologies with management strategies (in an integrated manner) to improve the safety and efficiency of the surface transportation system. ITSs also are advanced applications which without representing intelligence as such, aim to provide innovative services relating to different modes of transportation and traffic management in order to enable various users to be better informed and drive safer, more coordinated, and 'smarter' use of transport networks [1]. EU Directive 2010/40/EU on their own part defines ITS as a systems where information and communication technologies are applied in

the field of road transportation, including its infrastructure, vehicles, users, traffic management, mobility management, as well as the interfaces with other modes of transport [3].

ITS also is one of the most important implementations for analyzing and handling traffic control management of moving vehicles on roads in urban and mega cities. Combined automated traffic management systems are now being implemented across different cities in developed countries and other developing countries [1], [3]. The main objective of such system is to track down vehicles that violated traffic regulations using surveillance cameras and intelligent image analytic software. Owing to growth in the number of road users, like other modern cities in developed countries, Nigeria needs an intelligent, reliable and efficient traffic management system to cope with the constantly increasing traffic violations and accidents on their major roads [4].

Vehicle License Plate Recognition systems, an integral part of ITS, is popular and has been studied extensively especially in developed countries such as USA [5-10], UK [6], South Korea, Germany [7], China [2], [8], Ireland, Japan, Australia, Sweden [3][9], Canada, Denmark [10][11][12], India [4][18-23], Malaysia [14], Iran [15][16][17], Iraq [18], Brazil [19] and also in some African countries like Tanzania [20], Egypt [21], etc. Ironically, while a distinctive feature of this system in traffic control management is being applied widely in most of these countries, a countable number is being restricted to a specific region or city in developing countries like Nigeria which has not witness any type of this system [4]. The only work carried out on this wise was that of [22] making use of Hidden Markov Model for character classification. Likewise, many intensive research studies in areas of vehicle license plate recognition have been conducted in other developed countries using different models, nevertheless; there is little recorded research studies conducted in Nigeria. This is due to the irregularity of standardization in the Nigerian vehicle license plates in the past (i.e., the pattern/format, standard, dimension and the layout of the license plates, policy making and regulations) [22].

Therefore, major big cities in Nigeria accommodating road crossing traffics are controlled by automatic colored lights signaling system for interpretation of three types of signals for traffic controlling operations. For instance, green light signals by passing a stopped vehicle, yellow light signals a moving vehicle to slow down their speed and be ready to stop or a preparing a stopped vehicle to go while the red light signals the vehicle to stop.

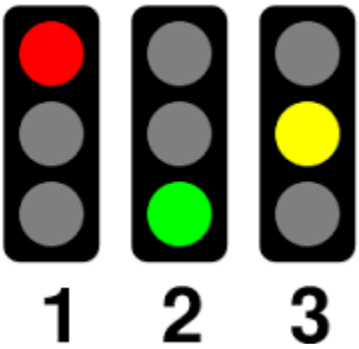


Fig. 1. Traffic Light Signals – Red, Green, Yellow [23]

Unfortunately, in many road crossing in Nigeria, it is not so occasionally. It is expected that when red light is shown on a lane, the signal conveys the message to all vehicles moving towards the crossing line to slow down and stop immediately. To make the procedure more systematic and convenient, a uniform thick white line is drawn on each lane before the road crossing. This line is commonly known as stop-line. Each vehicle coming towards the crossing must stop before this line if red signal light is seen by it. Even if the front wheel of the vehicle touches the stop-line partially then it is tagged a violating vehicle.

Stop-line is usually placed perpendicular to the direction flow of traffic and is placed in the plane of the road. Presently in Nigeria, the task for detecting the stop-line violating vehicles on a red light traffic signal is done manually by traffic police or the road safety agencies like Federal Road Safety Commission (FRSC). Unfortunately, most of the Nigerian road users do not comply with all these regulations. For instance, in a busy environment with hundreds of vehicles passing through the crossing it might not always be possible to manually generate a full list of all the violating vehicles. This is one of the primary motivations behind this paper work.

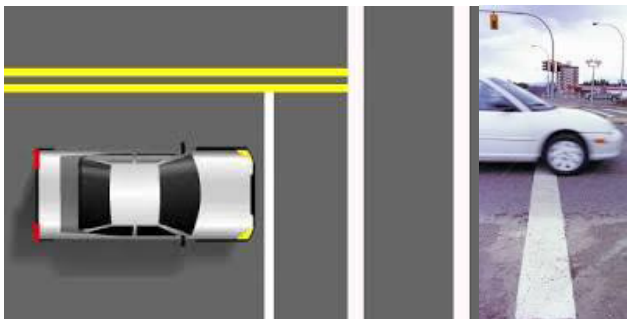


Fig. 2. Stop Line [23]

One of the effective solutions to control and manage traffic violation, previously being monitored by the human system (i.e. Police, Traffic Police, FRSC), is by employing an intelligent system that recognizes vehicle license plate numbers (pattern) using Pattern Recognition models. The paper's focus on intelligent system would identify violating vehicles, via various techniques which is going to be based on

algorithms rather than manual system procedures. Such algorithm comprising of different techniques will be using Pattern Recognition for this purpose. Image processing, one of pattern recognition techniques, deals with image sequences taken from vehicles at designated traffic lights points. All Nigerian vehicles (e.g., Diplomats, Federal, and State, Regulatory Agencies & Institutions, Private and Commercial vehicles) have their own vehicle license plate numbers which is a unique property and which takes into account for identifying vehicles. Quite a few monitoring functions can be supported by application of computer vision and pattern recognition techniques, such as the detection of traffic violations, traffic light violators, illegal turns, and illegal one-way defaulters through recognition of vehicle license plates numbers.

However, as part of the country's security measures, the government has mandated every citizen to be enrolled in the National Database under the supervision of the National Identity Management Commission (NIMC). The enrollment cut across every sector in order to curb crimes and violators of law and order. The electronic identity card issued by the NIMC bears the necessary information of individuals which makes it difficult for impersonation. The Federal Road Safety Commission (FRSC) and Vehicle Inspecting Officers who are the main factors in vehicle plate number issuance are in collaboration with the NIMC to have a robust database of the entire road users. The paper highlights will be making use of the Vehicle Inspecting Office and FRSC database with respect to the NIMC database of the Ekiti State Section Command as a case study.

II. PATTERN RECOGNITION

Patterns could be relatively defined as an entity that could be represented in various multiple occurrences such as fingerprint images, handwritten words, speech signals, DNA sequences, security bar codes, web page links, and human face identification [24]–[27]. Pattern Recognition (PR), on the other hand, is a branch of machine learning that focus on the recognition of patterns in a way to observe the environment or learn in distinguishing patterns of concentration by making sounds and reasonable decisions about the categories of patterns to be identified [23]–[32]. Humans always have a premonition and stimulus on the facts of how patterns are displayed and recognized in nature when procedures and methods of recognizing the pattern are developed. Different studies on machine discernment have also being a positive impact on better understanding and obligation for pattern recognition. In another perspective, models in pattern recognition are also applied to numerical studies which are not related to nature [25], [27], [28], [32].

Pattern Recognition schemes are implemented in so many ways, from the labeled trained data called *supervised learning* to the unlabeled data which can be solved or identify using other forms of algorithm to discover the heretofore unknown patterns called *unsupervised learning*. Different studies have been interwoven with Pattern Recognition. Studies like machine learning, data mining and knowledge discovery in databases (KDD), have been overlapping in scope with PR. Having an understanding of each of these studies briefly usually makes the process of image recognition easier and better. For clarity purpose, machine learning is the mutual term used for supervised learning methods (labeled trained data) which normally initiates from Artificial Intelligence (AI) while data mining and KDD, closely related in the business environment, have a bigger concentration on unsupervised methods (unlabeled data) [23], [32], [33]. In similarity, PR and machine learning have a close connection. In PR, there is a bigger attention and curiosity in enacting, elucidating and visualizing the pattern while on the other hand machine learning conventionally focuses on exploiting the pattern recognition rates [23].

PR can be categorized into various types of learning techniques used to produce the result value. As a general note, the supervised learning consisting of a set of occurrences assumes a set of training data that has been delivered and properly labeled with the correct output. Alternatively, unsupervised learning accepts training data that has not been hand-labeled like the supervised learning and endeavors to find the characteristic pattern in the data that can be used to determine actual result value for the new data occurrences [23].

Pattern Recognition of vehicle license plate numbers is carried out in the following manner

- Image acquisition (vehicle license plate localization),
- pre-processing of image (detecting the license plate),
- vehicle license plate extraction,
- character segmentation (vehicle license plate vertical and horizontal scanning),
- character recognition (recognizing the letters and numbers on the vehicle license plates), and
- vehicle license plate matching process (pattern template matching with the database)

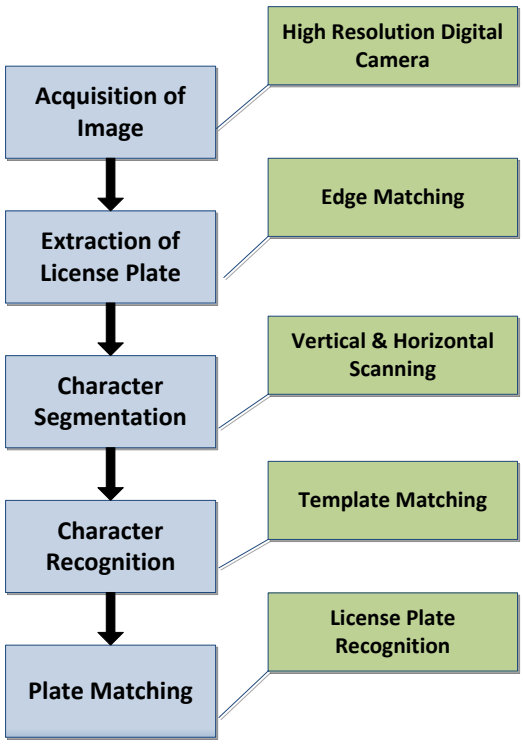


Fig. 3. Pattern Recognition Stages [6]

III. Study Motivation

As a result of technological advancement and increase in the number of road and vehicle users in Nigeria, there is need for an intelligent system to be in place to cope with the constantly increasing traffic light violations on Nigerian roads. The paper highlights using PR as a tool favors the recognition system for vehicle license number plate as a possibility of reducing traffic lights violation problems. The vehicle license plate recognition system should be accepted as mandatory due to too much increase of vehicles in Nigeria transportation system. Distractors factors such as sun light, shadow, mud, deleted plate characters, plate numbers ambiguity and irregularities and other factors in the vehicle license plate image make it difficult for existing system to recognize some plate recognition systems (i.e. the use of radio frequencies).

Apart from these factors, procedure applied for recognition also plays a vital role. This is necessary if the quality of the procedure is good, then more varieties of images can be given as input to the system, and this will also reduce the computational speed of the process. The basic issue in real-time recognition system is the accuracy, which this paper will address.

Nevertheless, it has been shown that there are umpteen setbacks parameters on which quality of image recognition depend which are the focus of this paper. Some of these hindrances are

- Image Quality – the image of the license plate to be captured must be of high quality in order to have good attributes for evaluation.
- Reduced tilting and spinning – the camera to capture the license plate image must be position at the right angle of view to guide against tilting and turning. The procedure to use in recognizing the image must also be accurate.
- Right Illumination – vehicle license plates are retro-reflective in nature, on this wise the camera illumination (flash) must be of high resolution to brightened the plates to be clearer in order to acquire the right attributes [6],[8], [9], [14], [19], [21], [34]-[38]

As part of the motivation, the paper demonstrates an effective framework using PR techniques as a tool for detection, extraction and recognition of vehicle license plate recognition system. The framework will work efficiently under all conditions of moving vehicles, adaptive to different traffic environment conditions; robustness against progressive or sudden illumination changes, occlusions, as well as image captured identification time of the system which should be as short as possible. The system should detect and recognize all the types of country's vehicles license plate numbers issued by the regulatory body (i.e. VIO as monitored by FRSC) to diplomats, federal, states, private and commercial plates. In addition, it should be resistant to any kinds of disturbances, which may occur in images and mechanical damages on the license plates which may appear in reality. The attributes of the vehicle license plates such as size, colour, font face size, colour of each character, characters' spacing, and characters' height and width, all play important role in the recognition process.

IV. Related Works

A considerably amount of studies have been carried out on image recognition of vehicle license plate numbers using pattern recognition as an intelligent tool but most of the research works (survey, test and results) are carried out in developed countries while few have been done in African communities. Few research works have addressed this vital study in Nigeria due to the unstable reforms in the vehicle license plate. The most related works to this paper are that of [34], [37], [38] where their focus is on real-time (moving) vehicles approaching the red light point. Each of the studies was able to capture, identify and recognize the characters of the captured vehicle license plate. Other studies like [1]–[3], [6]–[8], [10]–[22], [35]–[37], [39]–[55] used different pattern recognition methods to identify, capture, extract and recognize the vehicle plate number at still or packed environments. For instance [42] was a study that was carried out in the US using two algorithms – Integral Channel Features and Aggregate Channel Features detection models. The former has been used severally but the latter was a new technique employed by the researcher to make performance differences between the two techniques. Also in [14] which was carried out in Malaysia,

two algorithms were employed i.e. Connected component analysis and Perceptron Neural Networks (PNN), to recognize the characters of the vehicle license plates. The result shows that the techniques were effective for real time plate recognition. In addition, [8] and [20] were conducted in China and Tanzania respectively. In [8], statistical character method coupled with structural character method were used to obtain the character while fuzzy recognition method was later used for decision making while in [20], MATLAB algorithm was used for the character recognition. Another work for Malaysia was that of [50] where they used smearing algorithm to detect and identify vehicle license plates in still position in a large space. They believe that this algorithm can also be used for frameless vehicle license plates. [52] conducted a research on car license plates using neocognitron Artificial Neural Networks. In their work, they used an image processor, segment processor and combine it with five coupled neocognitron artificial type of neural network to perform their character recognizer. Their work was able to introduce the neocognitron neural networks for the purpose of removing the manual process of training the set of large numbers of ambiguous characters. They have success report of 94% using this type of neural networks for character recognizer. [1] on another platform focused on license plate recognition algorithm on ITS using optical character recognition and probabilistic neural network (PNN). The PNN model was trained to identify alphanumeric characters and their study had a success of 86% success but they believed they should have gotten a better result if the limitations of illumination and background complexity are taken care off. The only work that was carried out on Nigerian vehicle license plate was that of [22] where they used Hidden Markov Model (HMM) to perform the feature extraction and classification of the vehicle license plates. The work was a success at the time they carried it out but the only problem was that the policy and pattern of the country's vehicle license plates has changed and in that effect their proposed model can only be modify in order to be applicable for classification of Nigerian license plates at this time. Other mentioned related works used various algorithms from, ELMAN Neural Networks, Genetic Algorithms (GA), and profile method segmentation, etc. Each of the algorithms employed has played one or more important role in recognizing vehicle license plate number with recorded advantages and disadvantages

V. PROPOSED METHODOLOGY

A. Image Acqusition

Depending on the position of the vehicle license plate numbers, recognition systems normally consists of at least two high speed cameras (one high resolution digital camera and one infrared (IR) camera) to capture images of license plates. Furthermore, a software processor capable of performing sophisticated optical character recognition (OCR) to transform the image of the license plate into alphanumeric characters is needed. In addition, dedicated application software for comparing the transformed image characters to plate matching database of the monitoring agencies is highly recommended. Likewise, a user interface to display the images captured, the results of the optical character recognition (OCR) transformation, and an alert competence to notify operators when a violated vehicle license plate matched the monitoring agency's database/hotlist is observed is also recommended.

However, license plate number identification is an essential area in the development of intelligent traffic control systems. The usage of vehicles in Nigeria has increased rapidly due to high rate of urbanization and modernization, especially as Federal, States and Local Governments bodies are advancing, and thus, traffic violation is on the rise in big cities which has become major issues due to inadequate road infrastructure, lack of monitored and efficient policies and laxity of the existing human system. Therefore, control of vehicles and identification of traffic light violators to maintain order and discipline is becoming a big problem on most Nigerian roads. For this reason, development of an intelligent transportation regulatory system using the vehicle license plate number through patter recognition techniques to control traffic light violations is seen as a highly essential requirement at this time.

In developing countries like Nigeria, the attributes of vehicle license plates are strictly maintained. These attributes such as size of plate, color of plate, font face/size/color of each character, spacing between subsequent characters, background and foreground license plate color, etc.; are maintained very specifically. Figure 4 shows a sample of the approved Nigerian vehicle license plate, showing the various parts and their significance.

Some of the images of past and present standard vehicle license plates issued by the Vehicle Inspecting Office (VIO) controlled on the roads by Federal Road Safety Commission (FRSC) in Nigeria are shown in Figure 5.

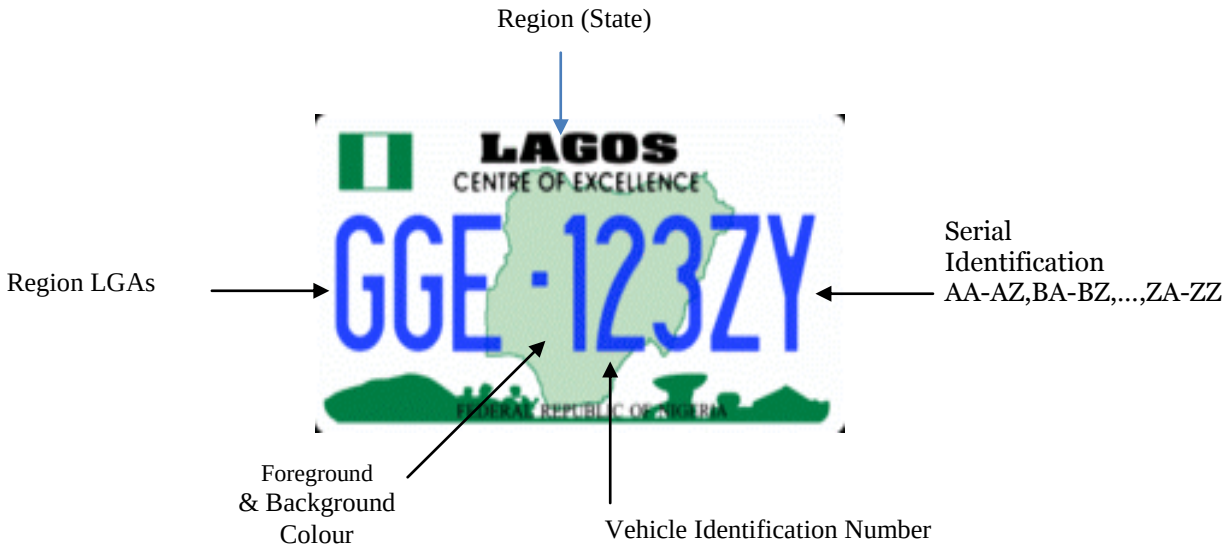


Fig. 4. Nigerian approved Vehicle License Plate Number with its features

B. *Plate Detection and Localization*

Image detection, recognition and traffic observations are important parts of intelligent transportation systems. Information about current situations of traffic violations can be automatically extracted by image processing techniques. Besides license plates detection, identification of vehicle license plate recognition with the regulatory agencies' database (VIO in support by FRSC) is very significant for this work. Different hindrances facing the processing of images captured from the vehicle license plates are numerous. Obstacles like poor image resolution; distance of license plates to the dedicated camera, low camera quality, motion blurring, poor lighting and low contrast due to overexposure, reflection or shadows, dirt on the plate, are covered within this work. Image enhancement techniques which is very crucial process based on filters is expected to be used to remove noise and unwanted light defects on the image to obtain clear and readable images; are all going to be evaluated in this work.

Figure 5 (a) and (b) shows Nigerian Vehicle License Plates numbers indicating approved and non-approved license plates; for both private and commercial road users. The distinct in the colours makes the feature segmentation and classification of the characters easier and the outcome of the result more successful.

Non-approved private license plates

Approved private license plates

Non-approved commercial license plates

Approved commercial license plates

Fig. 5. (a.) Private (b) Commercial Vehicle License Plates

Other samples of Nigeria vehicle license plate numbers are shown below in Figure 6 showing the trend of changes in the country's vehicle license plate numbers from 1976 till now for both private and commercial vehicle users.

Fig. 6. Trend of Nigerian Vehicle License Plate Numbers

In order to identify a violated vehicle by reading its license plate successfully, it is necessary to locate vehicle license plate image provided by issuing agencies (with respect to its characteristics – holder's designated body, characters & numbers) through some acquisition systems like a video or high-tech still surveillance camera. A number of commercial software in this area is readily available when images are provided in different styles and formats.

C. Image/Pattern Segmentation

One of the pronounce and efficient method that has been used over time in character segmentation is connected-component based method [3], [9], [34], [54]. The new Nigerian license plate numbers is designed in such a way that the features of the plate (i.e. digits and the characters) are in fixed position as well as lying in a horizontal orientation. In the past, license plate numbers can either be in horizontal or vertical orientation which can pose a problem in segmentation. In the event that the characters or digit of the license plate number image are not fixed or in a horizontal position, the image is rejected. For the purpose of this, horizontal projection algorithm is applied for character segmentation since the orientations of the license plates are in horizontal form and also it accommodates both horizontal and vertical positions.

D. Image Pre-Processing and Recognition

Image processing or character recognition techniques like edge-detection, thresh-holding, re-sampling and filtering could be used, as first starters, to locate the vehicle image and isolate the vehicle license plate and the characters from the vehicle image. These techniques alone will not be sufficient to meet the requirements of modern systems but can be coupled with other renowned techniques to achieve the right result. Intelligent vehicle license plate recognition system with reference to PR, as depicted in this work, is required to operate robustly in environments with complicated backgrounds and light intensity variations. Each of these techniques is focused on producing a result that is not of high computational cost.

A comprehensive well featured system for license plate image detection and recognition that would be obtainable would be available. Image detection and recognition system with applications in pattern recognition and machine vision which ranges from complex security systems to common areas from urban traffic control to semi-urban areas would be considered in this work. The system will cover complex characteristics of the image due to diverse hindrances effects such as fog, rain, shadows, uneven illumination conditions, occlusion, variable distances, velocity of car, scene's angle in frame, rotation of plate, number of vehicles on the scene and others.

E. Plate Matching

The rule of thumb for plate matching involves the process of employing template matching where it involves the use of a database (such as VIO or FRSC or NIMC) of characters or templates. The main objective of this work is to reveal a system that solves the practical problem of traffic violation on Nigerian roads using vehicle license plate identification for physical moving vehicles with emphasis on traffic light violations using pattern recognition techniques.

For proper license plate matching, separate template for each conceivable input character is created. In this sense, recognition is achieved by comparing the current input character with each of the database template for a possible match. For instance, if H(a,b) is the conceivable input character from the detection, and Z(a,b) is the template in the database of template *n*, then the template matching function say c(H,Z*n*) will return an output of the most possible likelihood of a match between template *n* and the conceivable input character.

The feature classes would then be compared with the target database of the issuing agency feature classes. The recognition structure would combine adaptive iterative thresh holding with a pattern/template matching procedure for the process of detecting a match with the database. The work is expected to be robust to illumination, character size and thickness, skewing and small character breaks. A plus to this work would be its competency in moving vehicles compare to previous works which focus on still or packed vehicles. With this aptitude, the system framework does not require any additional sensor input except the captured vehicle license plates image. This paper demonstrates that expected developed system would also be systematically

robust not only for this purpose but would also be efficient in suitability for other image recognition application. Furthermore, the paper highlights shows that the system does not need any installation on targeted vehicles, such as transmitter or responder for perfect recognition. Figure shows a block diagram of the methodology for recognizing vehicle license plate number.

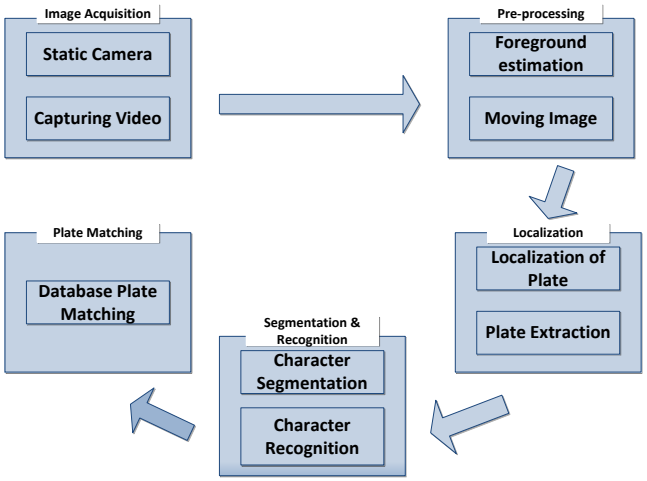


Fig. 7. Pattern Recognition System Block Diagram [38]

VI. Discussion

PR system is an ideal system that process patterns in different areas to handle real objects and noisy data as input into the system for recognition. Detection of moving violating vehicles is the first relevant step of information needed in recognizing the license plate of the violated vehicle. In addition, background subtraction is the ideal approach for foreground segmentation. Different background subtraction methods are needed in order to overcome the problem of illumination variation, shadows, background clutters and camouflage. Thereafter, Vehicle License plate extraction, an important stage in pattern recognition for intelligent transport system, should also be considered and this is a paramount aspect of the license plate recognition system. The extracted vehicle license plates would then be segmented into individual characters using a region-based method. For each character, feature classes would also be created in these order: from left to right, from right to left, from top to bottom and from bottom to top. Hence, PR system should be adequately accepted for image processing

VII. Recommendations

This paper recommends prompt implementation of this study by the government and public awareness and to the users of the road against traffic lights violations in Nigeria. Another point of reference is the positioning of the recognizing devices (cameras) must be placed in a position to only identify the license plate numbers. Also, the government is enjoined to enforce a law towards banners, posters, sign boards, etc., close to the image acquisition area so as to guide against conflicts between license plate numbers and other images. Reducing the rate of accidents on the road at traffic lights junction will reduce the high rate of accident victims in hospital and death as well.

VIII. Conclusion

Vehicle License Plate character recognition in Nigeria is paramount especially in the areas of traffic lights violation. PR system is the brain behind regulating intelligent transportation system in Nigeria. The paper has been able to demonstrate the importance of PR with respect to curbing traffic lights violations and as such reduce the high crime rates in Nigeria. The paper has also been able to show that combining different techniques can improve the positive output of plate matching. This combination can be done right from the detection stage after acquisition. Detection process or techniques that will be of good value but less computational cost is advised. Extraction process such as feature extraction and adaptive iterative thresh-holding can also be used for adequate template matching.

In another dimension, proper documentation on the part of the database template matching agencies is highly recommended. During the stage of vehicle user's biodata registration, appropriate names, address and other documentation must be done in order to guide against system failure during plate matching. This is necessary because of the previous history of nonchalant attitude to proper documentation among workers of these agencies due to laziness and not giving the work to the right expertise.

Acknowledgment

Transportation and Violation assistance was provided by the officers of the Federal Road Safety Commission and Vehicle Inspecting Office, Ekiti State Command and under proper direction and supervision of the authors. The authors received no funding from FRSC or VIO office.

References

[1] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, V. Loumos, and E. Kayafas, "A license plate-recognition algorithm for intelligent transportation system applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 377–391, 2006.

[2] Q. Fan, "Intelligent Vehicle Identification System based on Traffic Video Qiuyue," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 9, no. 8, pp. 417–424, 2016.

[3] E. Bergenudd, "Low-Cost Real-Time License Plate Recognition for a Vehicle PC," *Transform*, no. December, 2006.

[4] P. P. Randive, S. Ahivale, S. Bansod, S. Mohite, and S. Patil, "Automatic License Plate Recognition [Alpr] -a Review Paper," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 1, pp. 1100–1103, 2016.

[5] C. Li and C. Wang, "Intelligent Driving Situation Evaluation," *Traffic Eng. Control*, pp. 1–9, 2015.

[6] H. K. R. Medipally, "Licence plate localization and recognition algorithm using edge detection technique," *Dissertation*, pp. 1–55,

2010.
[7] M. Michael and M. Schlipsing, "Extending Traffic Light Recognition : Efficient Classification of Phase and Pictogram," *Inst. Neuroformatik*, pp. 1–8, 2015.
[8] L. Jin, H. Xian, J. Bie, Y. Sun, H. Hou, and Q. Niu, "License Plate Recognition Algorithm for Passenger Cars in Chinese Residential Areas," *Sensors*, vol. 12, no. 12, pp. 8355–8370, 2012.
[9] F. Trobro, "Real Time Automatic License Plate Recognition in Video Streams," *Test*, pp. 1–28, 2007.
[10] M. Bornø, M. Philip, B. Thomas, and M. Mohan, "Vision for Looking at Traffic Lights : Issues , Survey , and Perspectives," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–17, 2016.
[11] A. Mogelmose, T. B. Moeslund, and K. Nasrollahi, "Multimodal person re-identification using RGB-D sensors and a transient identification database," *2013 Int. Work. Biometrics Forensics, IWBF 2013*, pp. 1–5, 2013.
[12] M. P. Philipsen, M. B. Jensen, U. C. S. Diego, A. Møgelmose, and T. B. Moeslund, "Ongoing Work on Traffic Lights : Detection and Evaluation," *12th IEEE Int. Conf. Adv. Video Signal based Surveill.*, pp. 1–6, 2015.
[13] S. Saha, S. Basu, M. Nasipuri, and D. K. Basu, "Localization of License Plates from Surveillance Camera Images: A Color Feature Based ANN Approach," *Int. J. Comput. Appl.*, vol. 1, no. 23, pp. 27–31, 2010.
[14] O. Khalifa, S. Khan, R. Islam, and A. S. (International I. U. O. M. Kulliyyah, "Malaysia Vehicle License Plate Recognition," *Int. Arab J. Inf. Technol.*, vol. 4, no. 4, pp. 359–364, 2007.
[15] S. Rastegar, "An intelligent control system using an efficient License Plate Location and Recognition Approach," *Image Process.*, no. 5, pp. 252–264, 2007.
[16] S. H. Kasaei, S. M. Kasaei, and S. a Kasaei, "New Morphology Based Method for Robust Iranian Car Plate Detection and Recognition," *Int. J. Comput. Theory Eng.*, vol. 2, no. 2, pp. 264–268, 2010.
[17] S. H. Mohades Kasaei, S. M. Mohades Kasaei, and S. A. Monadjemi, "A novel morphological method for detection and recognition of vehicle license plates," *Am. J. Appl. Sci.*, vol. 6, no. 12, pp. 2066–2070, 2009.
[18] A. Mohsin, "An Automatic Recognizer for Iraqi License Plates Using ELMAN Neural Network," *J. Softw. Eng. Appl.*, vol. 3, no. December, pp. 1163–1166, 2010.
[19] P. R.F, C.-C. G, W. R. Schwartz, and M. D, "Brazilian License Plate Detection Using Histogram of Oriented Gradients and Sliding Windows," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 6, pp. 39–52, 2013.
[20] I. Bulugu, "Algorithm for License Plate Localization and Recognition for Tanzania Car Plate Numbers," *Int. J. Sci. Res.*, vol. 2, no. 5, pp. 12–16, 2013.
[21] M. M. Nabeel, M. Fakhrel-dein, and S. El-kader, "Intelligent vehicle recognition based on wireless sensor network," *Ijcsi*, vol. 10, no. 4, pp. 164–174, 2013.
[22] D. S. Adebayo, E. Adetiba, J. Badejo, A. J, Samuel, and T. Fagorusi, "Automatic Vehicle Identification System Using," *Int. J. Eng. Sci.*, vol. 3, no. 2, pp. 1712–1719, 2011.
[23] Wiki, "Pattern Recognition," *Wikipedia*, pp. 1–11, 2016.
[24] S. Aksoy, "Introduction to Pattern Recognition Part II," *Pattern Recognit.*, 2005.
[25] A. K. Jain and R. P. W. Duin, "Introduction to Pattern Recognition 1," *Pattern Recognit.*, pp. 1–5, 2004.
[26] S. Aksoy, "Introduction to Pattern Recognition Part I," *Pattern Recognit.*, pp. 1–26, 2005.
[27] S. Aksoy, "Introduction to Pattern Recognition Human Perception," *Pattern Recognit.*, pp. 1–40, 2016.
[28] J. Liu, J. Sun, and S. Wang, "Pattern Recognition : An overview," *Int. J. Comput. Sci. Netw. Secur.*, vol. 6, no. 6, pp. 57–61, 2006.
[29] S. Asht, R. Dass, A. . Fallis, T. Y. Dn, T. Yyepg, E. Cambria, G.-B. Huang, L. L. C. Kasun, H. Zhou, C. M. Vong, J. Lin, J. Yin, Z. Cai, Q. Liu, K. Li, V. C. M. Leung, L. Feng, Y.-S. Ong, M.-H. Lim, A. Akusok, A. Lendasse, F. Corona, R. Nian, Y. Miche, P. Gastaldo, R. Zunino, S. Decherchi, X. Yang, K. Mao, B.-S. Oh, J. Jeon, K.-A. Toh, A. B. J. Teoh, J. Kim, H. Yu, Y. Chen, and J. Liu, "Pattern Recognition Techniques: A Review," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2013.
[30] T. Slivka, "Pattern Recognition," *Pattern Recognit. Lett.*, no. November, pp. 1–11, 2010.
[31] K.-S. Fu and A. Rosenfeld, "Pattern recognition.," *Ieee Trans. Comput.*, vol. C-25, no. 12, p. 1336, 2006.
[32] N. J. Nilsson, "Introduction to Machine Learning," *Mach. Learn.*, vol. 56, no. 2, pp. 387–99, 2005.
[33] W. Is and I. Good, "Image Analysis and Pattern Recognition 1," *Canada Cent. Remote Sens.*, pp. 1–7, 2015.
[34] M. Kumar, "A Real-Time Vehicle License Plate Recognition (LPR) System," *Dissertation*, no. July, p. 77, 2009.
[35] L. Angeline, H. K. Lau, B. K. Ghosh, H. H. Goh, and K. T. Kin Teo, "Development of a License Plate Recognition System for a Non-ideal Environment," *Int. J. Simul. Syst. Sci. Technol.*, vol. 13, no. 3 C, pp. 26–33, 2012.
[36] David João Adão dos Santos, "Automatic Vehicle Recognition System- An approach using car rear views and backlights shape," *Dissertation*, no. September, 2008.
[37] S. Saha, S. Basu, M. Nasipuri, and D. K. Basu, "Development of an automated Red Light Violation Detection System (RLVDS) for Indian vehicles," *Natl. Conf. Comput. Commun. Syst.*, no. March 2016, pp. 59–64, 2009.
[38] L. Kodwani, "Automatic Vehicle Detection, Tracking and Recognition of License Plate in Real Time Videos," *Master Thesis*, 2013.
[39] S. Ozbay and E. Ercelebi, "Automatic vehicle identification by plate recognition," *World Acad. Sci. Eng. ...*, pp. 222–225, 2005.
[40] G. Garg and A. Kaur, "Study of Various Vehicle Detection Techniques –A Review," *Int. J. Adv. Res. Ideas Innov. Technol.*, vol. 2, no. 3, pp. 1–5, 2016.
[41] E. C. Sharma and E. A. Singh, "An Efficient Approach for Detection and Extraction of Vehicle License Plates using Edge Detection Technique.," *Int. J. Comput. Sci.*, vol. 5, no. 1, pp. 33–36, 2014.
[42] A. Mogelmose, D. Liu, and M. M. Trivedi, "Detection of U.S. Traffic Signs," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3116–3125, 2015.
[43] A. Mogelmose, M. M. Trivedi, T. B. Moeslund, Ieee, and A. Møgelmose, "Traffic Sign Detection and Analysis: Recent Studies and Emerging Trends," *2012 15th Int. Ieee Conf. Intell. Transp. Syst.*, pp. 1310–1314, 2012.
[44] M. S. Kristoffersen, J. V Dueholm, R. K. Satzoda, M. M. Trivedi, A. Møgelmose, and T. B. Moeslund, "Towards Semantic Understanding of Surrounding Vehicular Maneuvers : A Panoramic Vision-Based Framework for Real-World Highway Studies," *Traffic Technol. Int.*, pp. 1–8, 2014.
[45] S. Hosseinyalmdary, A. Yilmaz, C. Section, and H. Transformation, "Trafic Lght Detecttion Using Conic Section Geometry," *Photogramm. Comput. Vis. Lab.*, pp. 1–10, 2011.
[46] M. L. Ginsberg and G. I. Road, "Traffic Signals and Autonomous Vehicles: Vision-bsed or a V2I Approach?," pp. 1–11, 2015.
[47] A. Prioletti, A. Mogelmose, P. Grisleri, M. M. Trivedi, A. Broggi, and T. B. Moeslund, "Part-based pedestrian detection and feature-based tracking for driver assistance: Real-time, robust algorithms, and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1346–1359, 2013.
[48] M. P. Philipsen, M. B. Jensen, A. Møgelmose, T. B. Moeslund, and M. M. Trivedi, "Learning Based Traffic Light Detection: Evaluation on Challenging Dataset," *18th Intell. Transp. Syst. Conf.*, no. Submitted, pp. 1–5, 2015.
[49] M. P. Philipsen, M. B. Jensen, R. K. Satzoda, M. M. Trivedi, A. Møgelmose, and T. B. Moeslund, "Day and Night-Time Drive Analysis using Stereo Vision for Naturalistic Driving Studies," *IEEE Intell. Veh. Symp.*, vol. 2015, no. June, pp. 1–6, 2015.
[50] L. Angeline, K. T. . Teo, and F. Wong, "Smearing Algorithm for Vehicle Parking Management System," *Proc. 2nd Semin. Eng. Inf. Technol.*, no. July, pp. 331–337, 2009.
[51] a. Akoum, "A New Algorithmic Approach for Detection and

Identification of Vehicle Plate Numbers," *J. Softw. Eng. Appl.*, vol. 3, no. 2, pp. 99–108, 2010.
[52] H. Ri, H. Z. Qhxudo, Q. Lpdjh, S. Olfhqvh, S. Uhfrjql, and H. U. V Vwhp, "Recognition of Car License Plates unsin a neocognitron type of Artificial Neural Netwokr," pp. 1–12, 2015.
[53] P. Vijayalakshmi, "Design of Algorithm for Vehicle Identification by Number Plate Recognition," *Engineering*, 2012.
[54] M. A. Joarder, K. Mahmud, T. Ahmed, M. Kawser, and B. Ahamed, "Bangla Automatic Number Plate Recognition System using Artificial Neural Network," *Asian Trans. Sci. Technol. (ATST*, vol. 2, no. 1, pp. 1–10, 2012.
[55] K. Bora, "Car Plate Recognition," *Dissertation*, no. June, 2009.

Health Information Management System

(A Security Review)

Olayinka A. Yusuf
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
yyvogue@yahoo.co.uk

Jimoh G. Rasheed
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
Jimoh_rasheed@yahoo.com

***Abstract*—The call for adaptation of Health Information Management System in Hospitals and health care centers has rapidly grown over the past few years. Developed Countries have begun full utilization and deployment of Health Information Management Systems across the health sectors nation wide, while a few hospitals in Nigeria have started implementation partially with digitization of patient records with limited data like Names, age and house address. With the Nation wide implementation of the National Health Insurance Scheme (NHIS), adaptation of Health Information Management System has become a critical and vital necessity to enable patient receive adequate medical attention regardless of their location. The Integrity, Confidentiality and Availability of patient records are of immense importance for efficient usage of a Health Information Management System. This study reviews the need for full adoption and implementation of a Health Information Management System across board in hospitals in Nigerian to enable patient enjoy the benefit of the National Health Insurance Scheme in comparism to the present manual record keeping method on ground, while evaluating the security requirements of a Health Information Management System.**

Keywords—Health Information Management System; Patient Records; Cryptography and Security Requirements

I. INTRODUCTION

Health Care Sector is a very critical aspect of every economy and its importance cannot be over emphasized. Almost every human at a point in its life span will need medical treatment. From the young to the old, male, female and even animals also get health treatment. This replicates the large numbers of Hospitals, Clinic and health care centers we have in the Country, with each politically demarcated ward having at least a clinic or health care Centre within its boundary. However, these health care facilities are not sufficient to cater efficiently for the over 140 million Nigerians. As we have limited Health Professionals and lack Medical utensils and machinery in the health care centers.

The National Health Insurance Scheme (NHIS) was established in 1999 to provide affordable health care services for every Nigerian [4]. This signifies the importance of the health care sector and the need to make it accessible to every one. There are a limited number of medical practitioners in comparism to the number of patients. Hence the rotation of duties by medical staff, this makes it unrealistic to have a particular medical officer attend to a specific patient at all times. Hiring and firing of medical personnel, relocation of patient from one location to another, billing and other essentials makes the issue of record keeping very vital in the health sector. The common way health records are kept to by opening a file folder for every patient, which we are all used to in our various hospitals and clinics.

The manual method of record keeping via file folders has made a lot of people lose confidence in the Information Management of Health Records across board, as it is usually marred with loss of patient file, at times it takes a whole day before a patient file is found in the records department, missing contents in the file like drug prescription and bills record to mention a few. All these point towards information security risks.

The idea of Health Information Management System is to help overcome the information flow and sharing challenges, likewise to reduce the information security risks involved in the conventional method of record keeping, this has sprung up lots of researches, analysis, experiments, designs and implementation of various Health Information management systems in a bid to resolve some of the existing issues with Manual process. Health Information system has been in existence for over 3 decades now [2]. Deployment of security services and security infrastructures are key to the successful design, implementation and management of Health Information Management Systems.

Recently, there as been an increased agitation for the use of Health Information Management System across hospitals national wide to help reap some of the intended benefits of the NHIS, hence the need for a review and understanding of current trends and development of Health Information Management Systems.

II. HEALTH RECORDS

Every individual has a unique health/medical record and it's important for these data to be accurate at every point in time. This so critical such that a single change in a patients health record can lead to wrong drug prescription, wrong diagnosis, delay in getting treatment and even loss of life. A review of key health record data is necessary to understand the uniqueness and importance of keeping an accurate health record of a patient.

A. Patient Bio Data

The first point of contact for a patient at the clinic or hospital is the registration office. Bio data of patient are collected and the patient is duly registered at the health facility. Data like Name, Address, age, next of kin, phone numbers are all collected. This form the basis of patients record with the hospital and a file is subsequent open for the patient

B. Medical Records

Medical records of patient are filed in a single or multiple folders as the case maybe for ease of reference. All related documents to a patient are kept together. Due to the large number of patients that visit the hospital daily, the storage process of the files are untidy and lots of documents fall off or at times a patient's folder cant be located and this affects on-going treatments and medications already prescribed for the patient, as well as the medical history of the patient. This gives and patient and the doctors unnecessary challenges and at time is very costly as it results in death or delay in getting medical attention.

C. Health Facility Record

The health facilities has its own records, be it an hospital or clinic, they keep record of the staff, payroll, Store items, Drug log, lap results, departments/units and programme for the day to day running of the health center. Different department keep different data, it's always a challenge for management to collate and update these data centrally on a timely basis due to the manual method behind operated.

III. HEALTH INFORMATION MANAGEMENT SYSTEM

Health Information Management system is a whole-listic approach in the management of health information/data management at a health facility center. The approach taps into the resourceful usage of ICT in solving challenges faced with the manual and untidy way of doing things into a more accepted electronic means inline with global world best practice.

Wikipedia defines Health Information Management System as "information management applied to health and health care. It is the practice of acquiring, analyzing and protecting digital and traditional medical information vital to providing quality patient care "[1] and it can also be defined as An Information System specially designed to assist in the management and planning of health programme, as opposed to delivery of care [3].

Health Information Management System comes in various forms and designs aimed at specific aspect in a health facility center, but regardless of the aspect aimed at the Patient/Doctor relationship must be taking to consideration because that is the basis for the existence of a Health Facility. Patient's Bio data and medical record are always critical, Most Information systems proposed or designed are to help, diagnose, experiment, report, maintain, collate, view or register patients to the Health Facility.

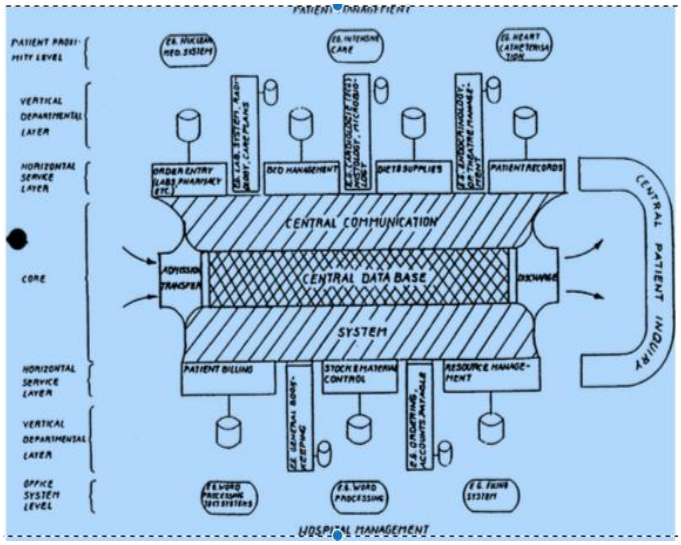


Fig. 1. Model of Hospital Management System

The figure above shows a model design for Hospital Management System [2]. The design of a health Information Management System should adopt a well laid out System Development Life Cycle (SDLC), as to the penetrate and patch method adopted by most software developers [10]. The stakes are very high when patients health record are involved. Saving of life is much more important than making money. As the Health Information Management System is expected to focus more on saving life and not a revenue generating item.

IV. SECURITY REQUIREMENTS

Security requirement are very key in design and implementation of a Health Information Management System. Security in systems is usually evaluated by how much they hold with regards to the core information security tenets with are Confidentiality, Integrity and Availability usually referred to as CIA Triad. Hence a need to review this core security tenets.

A. *Confidentiality*

Confidentiality offers a high assurance that data, information, objects or resources are restricted from unauthorized subjects [6]. With regards to Health Information, data like patient medical records require high level of privacy and confidentiality, such that the content of a patient's medical history is not disclosed to unauthorized people. This ensures that patient's medical history is kept discrete and private.

Even in advance countries like the United States of America and United Kingdom data theft in the health sector is rampant, with such data been sold on the dark net and privacy of top figures in the society been exposed. This often leads of lawsuits or monetary settlements.

Some of the counter measures for containing confidentiality include; Cryptographic encryption, Data classification, Strong access control and authentication measures [6]. A good Health Information Management system must take into consideration these measure from the design inception. As security is best implemented from design stage rather than an after thought.

Fig. 2. CIA Triad

B. *Integrity*

Integrity is the tenet of information security that assures that data is not altered mistakenly or intentionally by an unauthorized person while the data is at rest (storage), in transit or during processing i.e. Changes on data can only be done by authorized personnel [7]. Integrity is highly important in a health care facility data; hence a Health Information Management system must maintain a high level of integrity, Patients record must not be altered intentionally or mistakenly at any stage/time. Results, diagnosis, prescription and so on must be kept intact. If an unauthorized alteration occurs it should be detected. This can be mitigated cryptographically with integrity checks like cyclic redundancy check (CRC) and hash functions. Likewise strong access control measures [15].

C. *Availability*

Availability gives the assurance that subjects are able to access objects (data) at every point they need to [6]. Patients Bio data and medical record at a health facility must be available at all time its needed to enable doctors attend to patients. It's critical for a health Information management system to have a back up plan (Business continuity plan/ disaster recovery) amongst other mitigation methods [7]. There are other security tenets that should be considered like non-repudiation, audit and authentication [15]. It should be stated again that these core security tenet should be integrated to the system from the design stage. The use of Cryptographic Primitives will be very resourceful in implementation the core security tenets like Public Key Infrastructure (PKI), since a lot of information flow is involved which needs to be protected at rest, in transit and storage [12,13].

V. NETWORK AND COMUNICATION

The use of computer and mechanically equipment has come of age in the Health Sector, such facilities are used for record keeping, diagnoses, experiments, payment transactions, information gathering and sharing [11] to mention a few. Computers are only as secured as the security implemented by the user and more secured as a stand alone [8], but with sharing of information, more security measures must be taking. The design and implementation of a health information management system cannot be completed without a medium for data flow/sharing secret information [13]. This is usually achieved by the use of network and communication devices. The use of Network and communication devices comes with its advantages as well as flaws that should be mitigated against. This flaw includes but not limited to Denial of Service (DoS), Single Point of Failure, Hackers, Virus, Malwares and so on [9].

Payment of bills is another major challenge and this aspect is also integrated into the Health Information Management System and allows patients to settle bills electronically, also receipts are issued electronically and updated immediately in the database. The security of the web payment platform must be top notch to avoid online attacks, stealing of patient's financial details and unauthorized transfer of funds [14]. The use of personal electronic devices by staff is another dilemma security wise for the management of a Health facility center about to implement a Health Information Management System.

VI. CONCLUSION

The idea for the design and implementation of Health Information Management System across hospital and health care centers is good and should be encourage, however its very important that security of data at various stages are considered. Design of Health Information Management Systems should adopt a well laid out development life cycle, to ensure the system is robust enough and security is an integral part from conception. Testing of the system before

deployment should be done rigorously. Even if a system is designed and implemented with a good security structure, the system will only be a secured as the personnel using it. When is comes of security of systems, humans are always the weakest link. Hence it's recommended that every user is adequately trained and retrained.

Integration of other component or systems into the Health Information Management System should also be scrutinized critically and back up plan should be operational and tested from time to time. Future work would model a Health Information Management System suitable operationally in the Nigeria Health Sector environment.

References

[1] https://en.wikipedia.org/wiki/health.

[2] Haux, Reinhold. Health information systems–past, present, future. International journal of medical informatics: pp268-281. 2006.

[3] World Health Organization. "Developing health management information systems: a practical guide for developing countries." (2004).

[4] http://www.nhis.gov.ng/home.

[5] Lippeveld, Theo, Rainer Sauerborn, and Claude Bodart, eds. *Design and implementation of health information systems*. Geneva: World Health Organization, 2000.

[6] Harris, Shon. *CISSP all-in-one exam guide*. McGraw-Hill, Inc., 2010.

[7] Tittle, Ed, James M. Stewart, and Mike Chapple. *CISSP: Certified information systems security professional study guide*. John Wiley & Sons, 2006.

[8] Stallings, William, and Lawrie Brown. "Computer security." *Principles and Practice* (2008).

[9] Larry L Peterson and Bruce S Davie. Computer networks: a system approach. Communications Magazine, IEEE, 36(5):54–54, 1998.

[10] Jacobson, Ivar, et al. *The unified software development process*. Vol. 1. Reading: Addison-wesley, 1999.

[11] Adi Shamir. How to share a secret. Communications of the ACM, 22(11):612– 613, 1979.

[12] Matt Blaze, Gerrit Bleumer, and Martin Strauss. Divertible protocols and atomic proxy cryptography. In Advances in CryptologyEUROCRYPT'98, pages 127–144. Springer, 1998.

[13] Whitfield Diffie and Martin Hellman. New directions in cryptography. Infor- mation Theory, IEEE Transactions on, 22(6):644–654, 1976.

[14] Keith E Mayes and Konstantinos Markantonakis. Smart cards, tokens, security and applications. Springer, 2008.

[15] Keith M Martin. Everyday Cryptography: Fundamental Principles and Appli- cations. Oxford University Press, 2012.

A Novel Approach to Preventing Buffer Overflow Attacks Using NIDPS

Benjamin Aruwa Gyunka
The Directorate of Information and Communication Technology
National Open University of Nigeria (NOUN)
Abuja, Nigeria
gyunkson@gmail.com

Abikoye Oluwakemi Christiana
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
Kemi_adeoye@yahoo.com

***Abstract*— Buffer overflow are dangerous and mostly targeted network attacks against both mobile devices and desktops computers which are caused by software application coding vulnerabilities. Software vulnerabilities led attacks are critical and have very expensive impacts given that most organizations, and particularly the healthcare industry, have now resorted to these technologies for better services delivery. Software applications are developed these days to handle almost any operations in all sectors such as banking, commerce, healthcare, transport, schools and lots more. This demand has led to an increased number of software applications developed on daily basis resulting to the existence of lots of coding errors and security bugs in today's software applications. Applications created in C and C++ programming languages are more susceptible to buffer overflow attacks. This work researched into the different approaches available for the control and protection against buffer overflow attacks. Firstly, the host-based intrusion detection and prevention system (HIDPS) was analyzed. This defense system includes the Canaries, data execution and prevention (DEP) and address space layout randomization (ASLR). Secondly, network-based intrusion detection and prevention system (NIDPS) was also critically analyzed in which case the working of snort was closely examined. The analysis of the different security measures reveals that no one unique technique is capable of providing an all round protection. Each has its strength and weaknesses. The research therefore concludes by proposing a defense system that involves combinatory security techniques.**

Keywords*— *Buffer-Overflow, Stack memory, Heap Memory, Signature-based, Anomaly-based, NIDPS, HIDPS, ASLR, DEP, Network

VII. INTRODUCTION

Buffer is another technical term for a system's memory. Every memory block or buffer has a specific quantity of data or string length it is meant to accommodate each time. When a larger amount of data goes into this memory space, the space will not be able to contain it so the extra data is forced to overwrite data values which are in memories addresses that are adjacent to the original specified destination. This process is called buffer overflow or overflow of designated memory space. Overwriting adjacent memories addresses values can only be stopped by programs or applications that have strong bound checks which are able to stop and remove such excess data that are sent to memory buffer. The effect of buffer overflow attack is highly profound. A successful buffer overflow attack can result to memory access error, system crash, incorrect results, or even give the attacker the unrestricted privilege to modify system's internal variables [1] and [2]. Buffer overflow is one of the leading attacks against software and mobile applications asides SQL Injection which are both caused by poor software design and testing [3] and [4]. The attack occurs mostly in software that contains vulnerabilities such as programming errors and can cause irregular and dangerous program behaviour [2]. Two types of buffer overflow attacks exist, stack-based overflow and heap-based overflow, and their characterization is based on the locations of the buffer in the process memory [5].

Software and application developers, for market share purposes, rapidly develop these applications software at a very high rate such that, most times, less security attention is usually given to the coding, so serious security bugs exist in them. Not all buffer overflow attacks are deliberate; while some of the attack incidences just happened coincidentally because of the bugs existing in the software; most others are deliberately crafted and targeted by malicious actors who sent data larger than the designated memory space. The attack is mainly targeted at overwriting the function's return pointer – it is a pointer found at the adjacent memory whose responsibility is to point the process to the next address to go after inputs have been taken, it is also known as the Extended Instruction Pointer (EIP) [1]. Once the attacker succeeds at exploiting and overwriting this function return pointer, he is then able to change the return function values in order to point to a destination that will fulfill his malicious intents. The original process execution path is then altered by this attack and thus complete control is transferred to the attacker's malicious code causing a program to execute something different from its original design, even when running with root privileges [6].

Buffer Overflow attacks varies depending on operating systems and programming languages. It is basically an exploit caused by improper input validation and can succeed in almost every programming environment that allows the manipulations of memory directly [1]. The attack vulnerabilities are more prevalent among the C and C++ programming platforms because they do not have an inbuilt security capacity to enforce bound checks, so they are unable to protect against accessing and overwriting of data outside an allocated buffer. These two programming languages also allow for direct memory manipulation through common programming constructs. Other programming environments like C#, Java and Perl are considered modern programming languages and they have reduced risks of coding errors and thus reduced buffer overflow vulnerabilities in software created through them [1]. Memory manipulation is usually made possible through errors in the software compiler, runtime libraries or features of the language itself. The Android mobile platform is also under buffer overflow threats with particular target on the Android software development kits (sdk) [7]. This exploit on the Android platform affected all versions of the Android debug Bridge (adb) which is on Linux x86_64.

This work seeks to explore different protection and remediation approaches available for buffer overflow attacks. Particularly, the study focuses on the different security provided by host-based intrusion detection and prevention systems (HIDPS) and the network-based intrusion detection and prevention systems (NIDPS). The main aim of the study is to carefully and critically analyze each one of these different security measures, analyzing their strengths and weaknesses, in order to, at the end, come up with a proposed defense system that would be capable of providing a more robust protection against this dangerous threat.

VIII. HOST-BASED INTRUSION DETECTION AND PREVENTION SYSTEM (HIDPS)

Host-based defense mechanisms are usually inbuilt in the given system. Mostly they come as software packages ready to be installed to provide the necessary security to the operating system kernel and to also monitoring and keeping track of every activities happening within the system [8]. These internally built security mechanisms include the Canaries, data execution and prevention (DEP) and address space layout randomization (ASLR). The host-based security techniques are deployed in order to protect against control flow hijack exploit, particularly the buffer overflow attacks [9]. At the application level, the SigFree is another invaluable host-based technology that provides both host and application security against buffer overflow attacks. The effectiveness and limitations of each one of these host-based technologies are considered in the sub-sections below.

A. The Functionality of Data Execution Prevention (DEP)

Data Execution Prevention (DEP) and Address Space Layout Randomization (ASLR) are two most profound and inter-twined security defenses inbuilt into operating systems such as Linux, Mac OS X, FreeBSD, and Microsoft Windows that guard these OS against both traditional and classic buffer overflow attacks. The data execution prevention (DEP) works by creating a harder environment that makes it extremely difficult for the injection and execution of any new and untrusted code by any attacker and thus preventing the successful execution of any shellcode [9]. The operational ideology of DEP is anchored on the fact that either memory should be writable or executable but must not take up both positions at program runtime. Certain memory structures, such as the stack, are usually forced to be marked as non-executable by the DEP [10]. For attacks that depends on using shellcode execution are mostly not successful in any environment where DEP is activated.

1) The Strength of Data Execution Prevention (DEP): The free and open execution access of any code in any environment must be carefully supervised as majority of successful cyber attacks are as a result of malicious code execution. Data execution prevention (DEP) provides a harden platform that stops the injection of new malicious code by any attacker into the system, therefore blocking new buffer overflow attacks.

2) The Short-comings of Data Execution Prevention (DEP): As it is with most security mechanisms, attackers are always looking for ways to circumvent these defenses and bypass detection. Same is true about DEP; it is unable to protect against JIT (just-in-time) compilers programs. These programs deployed a manipulative approach by changing their appearances to bypass DEP because DEP is implemented in such a way that only protect code modules that are explicitly labeled as DEP-safe [9]. Code reuse is another attack techniques used to bypass DEP. Since DEP fight and protect strongly against the injection of new code, a devised means of using code already found in the system to launch an attack always proof successful as the DEP system does not give any much careful attention to validate the code again before executing them. DEP focuses attention mainly to protecting the stack by preventing new codes injection and execution but is unable to protect the system against general buffer overflow and heap overflow. DEP does not address Return-to-libc attacks which is a kind of exploits caused when overflow modifies the return address and replaces it with the address of libc function such as "/bin/sh".

B. The Working of Address Space Layout Randomization (ASLR)

ASLR is not a technique that removes security vulnerabilities away from a system but it rather deploys a method of obfuscating these vulnerabilities via

randomization of some basic components of the memory in order to make it extremely difficult or nearly impossible for an attacker to either spot or exploit them [5] and [10]. The randomized components include all dynamic library addresses, stack and heap memories, and the code segments of executable of the memory spaces at loads and link time [11] and [12]. The obfuscation method deployed by ASLR makes it difficult or impossible for an attacker to know where any resources are found in the system. So in case of any compromise in which the overflow overwrites the extended instruction pointer (EIP), the attacker is prevented from setting the new pointer value to target the stack memory because he would not know what direction to take. The more bits of the operating system that are available to be randomized, and the speed of the randomization defines the strength of the defence the ASLR can offer. The greater the number of bits used the more a stronger and a reliable defence ASLR provides. Exploits such as buffer overflow attacks largely depends on the static nature of memory address values in order to succeed therefore the randomization is particularly focused on these static address values [10]. Three major factors determine the elasticity of ASLR defense. They are as follows [11] and [5]: (i) The predictability of the random memory layout of a program (ii) The ability of an exploit technique to tolerate variation in memory layout (iii) The number of exploit attempts an attacker can practically make.

1) The Strength of Address Space Layout Randomization (ASLR): This technique covers up the loophole of DEP by being able to protect against code reuse attacks and also can prevent successful execution of shellcode [9].

2) The Weaknesses of ASLR: As earlier stated, ASLR select only some memory components to randomize, this serves also as its weakness. The number of memory code ASLR is able to randomize also varies with operating systems. The unrandomized components, however small (as small as 20KB), are enough for an attack to take advantage of to launch very serious attacks like return-oriented programming and return-to-libc attacks [13] and [14]. Another deficiency with the ASLR is that randomization is done mostly for the base address of shared-library while living out many of the user level programs unprotected [15].

C. Canaries Methodology

Another name for canary is Stack-Guard. Compiling any application together with canary will automatically result to the application suddenly terminating each time traces of buffer overflow attacks are detected. Canary protect the stack memory against attacks, it does this by implementing an integrity verification of stack frames before function return address is called. The placement of the canary is made on the stack frames that are next to the return function. Any successful attack would overwrite the canary in a bit to overwrite the return function, once this is done, the canary gets corrupted and the system cease from work because it is designed to work with the canary. Values of canaries are very difficult to be corrected guessed by any attacker because these canary values are randomly generated with every new process [16]. This protective and decoy characteristics of canary makes it impossible for an attacker to overwrite the extended instruction pointer (EIP).

1) The Weakness of Canaries: The protection offer by cannries does not cover the entire memory space, it only protect the stack against overflow exploit and it also adds overhead. Though it's difficult guessing the correct canary values, however, NULL canary can be overwritten and Random canary can still be learned.

D. SigFree

The SigFree technology provides protections for web applications against buffer overflow attacks. It does not operate base on signature and thus provides real time protection. Just like the DEP, the SigFree also protect against injection and execution of codes. The operational ideology of the SigFree is based on the assumption that in client communication request sequence over the internet, executable code are usually not included [17] and [18], and so any request sequence that contains suspected executable code is strongly fought against by SigFree which then tries to remove the malicious code from the original request sequence. On spotting the malicious request, SigFree swiftly terminates the whole line of communication without given much attention to further analysis, it then sort out, in order of occurrence, the different commands or instructions of the particular request. When the entire sorting and separation processes are completed, SigFree moves further to make use of data flow anomaly to filter out all the undesirable instructions from the request sequence. On completion of the code extraction, SigFree then runs a comparison test in order to match the desirable request instruction sequence against a threshold to investigate the presence of any malicious code [19].

1) The Benefits of SigFree: SigFree does not rely on string matching and so it is immune against most attack-side obfuscation attacks. This technology is signature free and thus can identify and catch new and unknown buffer overflow attacks. The throughput degradation is very insignificant and the technology is less costly to deploy and maintain. SigFree makes use of generic code-data separation method rather than limited rules and [19] noted that SigFree has the strong ability to block all kinds of code injection attacks packets above 250. This technology has a low performance overhead.

2) Limitations of SigFree: Exploits that do not deploy code injection in corrupting control flow cannot be spotted by SigFree. For over flooded buffer found inside a JPEG or GIF system, ASN.1 or base64 encoder, SigFree cannot be deployed directly [19].

IX. NETWORK-BASED INTRUSION DETECTION AND PREVENTION SYSTEMS (NIDPS)

The internet network is largely exposed to different kinds of cyber nefarious activities, including buffer overflow attacks. Deploying a defence mechanism at the network layer is an essential decision to guard against such kinds of exploits. Network intrusion detection and prevention systems (NIDPS) are particularly designed to provide network layer security against known and unknown threats.

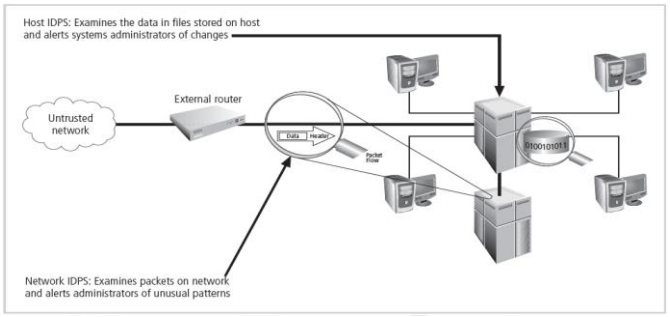


Figure 1: The operations and implementation of HIDPS and NIDPS [26].

Intrusion detection systems (IDS) are a mechanism that helps in monitoring intrusive and suspicious activities against security while intrusion prevention systems (IPS) on the other hand is an advanced system that comprises of both IDS, personal firewalls and anti-viruses put to function together in preventing malicious intrusive network attacks [20]. Network attacks usually exploit the vulnerabilities of network protocols and network-aware processes [2]. NIDPS evaluates traffic by looking for attack pattern and may be able to filter shellcode. NIDPS are of two types; signature-based and anomaly based NIDPSs.

A. Signature-Based Network Intrusion Detection System (NIDS)

Signature-based Network Intrusion Detection System (NIDS) works by matching the different network traffic with existing signatures of known attacks so as to monitor suspicious traffic in order to spot network attacks. The detection depends on pattern matching. Given that detection completely depends on the signature registered in the detection engine, any new and unknown attack whose signature is not found in that engine can never be detected by this system. This limits the detection capacity of this very system. The more frequent the database engine is updated with new attack signatures, the better chance the system stands to block more attacks. A good example of signature-based NIDS is Snort that can be deployed against buffer overflow attacks.

1) The Working nature of Snort: Snort functions by monitoring and analyzing network traffic and the flow of data in real-time. This ability enables snort to offer protection based also on the behaviour of network processes. It blocks any process or traffic whose behaviour does not appear normal with the other usual traffics even when such a process does not have its signature in the engine's database but will afterward log the details in the database [22]. This behaviour makes snort stand out in the ability to detect lots of suspicious packets of different types, particularly the buffer overflow attacks [8] and [22]. The snort engine comes with pre-defined rules or signatures but also provide allowance for new rules to be written into it. The components of snort include a preprocessor, output plug-in, detection engine, packet decoder, logging and alert systems. These components serve as extra extension to the functions snort can perform if well managed by the user [23]. The operations of Snort includes packet sniffing, packet logging and also serves as NIDPS [22].

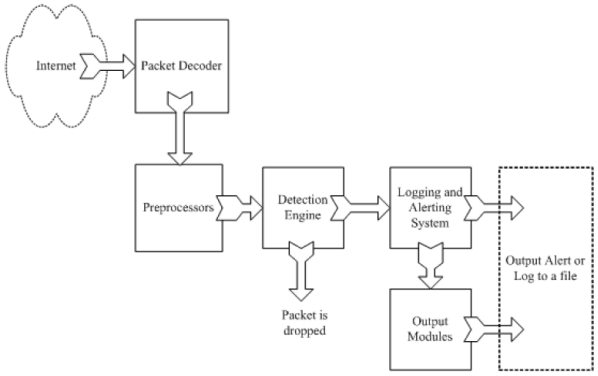


Figure 2: Components of Snort [19]

When a packet comes into snort, it goes into the preprocessor first. The preprocessor has the ability to execute and access the contents of packets to alter them by reassembly or rearranging their order. This enables snort to be able to block crafted and fragmented code strings, no matter how much an attacker tries to either manipulate or disguise then to circumvent existing signature-based detection. The preprocessor works by incepting such packets, breaks and rearranges them in order to reveal their original nature or identity, and then afterward the detection engine goes on to compare the packets with existing signatures and finally blocks them from passing through if found maliciously. The preprocessors adopts the principle of de-fragmentation, thus buffer overflow payloads that have been well crafted and fragmented may find it very difficult to escape snort detection.

2) Benefits of Signature-based NIDS: Any packet that come through this system and has its signatures registered in the detection engine can hardly pass through undetected. The system is also known to have very low false positive and it is very easy to implement.

3) Limitations of Signature-based NIDS: For attacks whose signatures are not in the registry of the detection engine, it becomes nearly impossible for the system to detect and block. Therefore, the system cannot protect against zero day attacks; they introduce maintenance

overhead, and are not very resilient to attack-side obfuscation [19]. The system is susceptible to compromise via polymorphism worms; because these worms are able to diversify their payloads for each infection attempt they make [25]. They are often too reactive.

B. Anomaly-based Network Intrusion Detection System

The anomaly-based system works by taking careful record of the state of every component in found in the network and in the computer system, it then stores and registers these records in its analysis engine. The components recorded may include kernel information; information of software running, system logs events, operating system information and network packet information. Any packet that comes through whose behaviour is different from the kind of behaviour recorded from any one of the initially recorded components, the system trigger and alert to indicate an intrusion and it then blocks the packet from passing through [8]. The technology can be deployed to guard against both known and unknown network threats.

1) Importance of Anomaly-based NIDS: Buffer overflow attacks can be strongly guarded against using this system especially that it can be used against both known and unknown threats. The system is able to analyze protocol against vulnerabilities and thus can protect against attacks before they occur [2].

2) Disadvantages of Anomaly-based NIDS: Because the system basically depends on behavioural activities to declare a packet suspicious, it therefore has very high level of false positive and false negative alarms. It also posses High overhead and processing capacity [21].

C. Comparisons between anomaly based and signature based NIDPS

Anomaly detection system can be used to provide protection against both new and old buffer overflow threats. On the other hand, the signature-based is largely dependent on only the registered known signatures its database to provide any protections. The signature-based system cannot protect against zero day attacks and it requires a constant updating of its database in order to offer a better and stronger protection. The anomaly based has a higher rate of false positive compare to the signature based system.

X. PROPOSED NOVEL APPROACH FOR BLOCKING BUFFER OVERFLOW ATTACKS

Having considered the different detection techniques for the detection and protection against buffer overflow attacks, it has been noticed that though most of these solutions have unique strengths, they can only provide a one sided protection. Most could be circumvented by one obfuscation methods or the other. Of all the different defense techniques analyzed, both the anomaly and signature-based network intrusion detection and prevention systems (NIDPS), if deployed to run simultaneously together would go a long way to reduce or even completely block the successes of buffer overflow attack on the network.

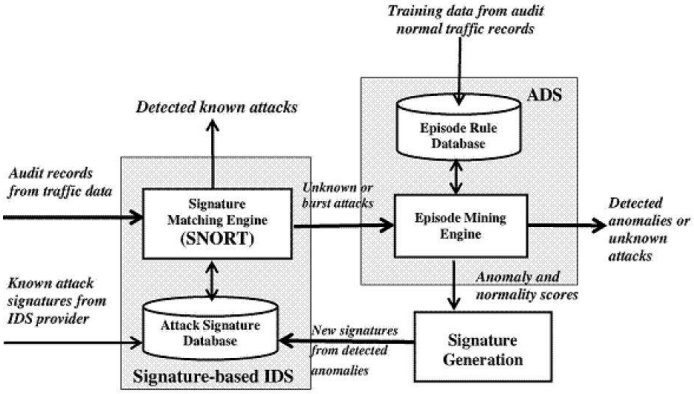


Figure 3: Signature-based & Anomaly-based combined IDPS Defense system

The two NIDPS systems would stand to complement each other in terms of their strength and weaknesses. Since one is able to have all the signatures of known attacks registered in its data, it would be able to protect against all the attacks that fall in that category. The other system is able to protect against both the known and the unknown attacks, although its rate of false positive and false negative is high. The de-fragmentation nature of snort will make it very difficult for most crafted attacks to pass through undetected. The signature based system will keep updating itself with new and most current signatures from all the block attacks. The location at which an IDPS is deployed is very crucial as it defines it characteristics and performance. NIDPS is deployed on the network, in a central location where its sensor is put at network ingress point to monitor all traffic passing through. With the implementation of the NIDPS on the network, buffer overflow will thus be strongly resisted. A stronger recommendation would be that both Host-based and Network-based Intrusion Detection and Prevention Systems (IDPS) deployed as a combined solution on the network.

XI. CONCLUSION

It is evident from this study that Buffer Overflow attacks can have very serious impacts on any sector they succeed. Successful attacks can result in server break-in, worms infections and propagation, zombies, and Botnets. Different approaches to completely block buffer overflow have been studied in this paper. The findings revealed that no single defense mechanisms can stand along to overcome the threats of buffer overflow attacks. This is because most of the security measures are only stronger at protecting against the attack from one point while living the other points vulnerable to the attacks. Recommendation from the various analysis would be that, in order to have a robust defense against this threats, combined security measures, host-based and network-based intrusion detection and prevention systems (IDPS) must be deployed concurrently. Another recommendation would be that application created in C or

C++ should keep away from using dangerous standard library functions, which lack the capacity to have bounds checks, such as gets, scanf and strcpy, but should rather explicitly use libraries or classes that are developed to perform string and other memory operations securely. Other security measures include patching of vulnerable software applications and adoption of runtime protection by different operating systems. This solution will be of immense advantage if deployed by powerful organizations such as the healthcare industries in protecting very important data. Having robust internet network and a secure Information and Communication Technology (ICT) infrastructure will enhance a protective information management culture and the dissemination of secure information to the public, particularly in the control of communicable and non-communicable diseases by the healthcare sectors.

REFERENCES

[1] M. Rouse, "Buffer Overflow," 17 August 2014. [Online]. Available: http://searchsecurity.techtarget.com/definition/buffer-overflow. [Accessed 30 October 2016].

[2] P. B. Pawar, N. Malti and P. K. Kawadkar, "Prevention of Buffer Overflow Attack Blocker Using IDS," *International Journal of Computer Science and Network*, vol. 1, no. 5, pp. 1-5, 2012.

[3] P. A. Williams and A. J. Woodward, "Cybersecurity vulnerabilities in medical devices: a complex environment and multifaceted problem," *Medical Devices: Evidence and Research*, vol. 8, p. 305—316, 2015.

[4] M. Bob, B. Mason, P. Alan and K. Dennis, "2011 CWE/SANS Top 25 Most Dangerous Software Errors," CWE/SANS, 2011.

[5] O. Whitehouse, "An Analysis of Address Space Layout Randomization on Windows Vista," *Symantec advanced threat research*, pp. 1-14, 2007.

[6] PicaTesHackZ, "An Introduction To Buffer Overflow Attack With Example," 18 June 2015. [Online]. Available: http://picateshackz.com/2015/06/an-introduction-to-buffer-overflow-attack-with-example.html. [Accessed 31 October 2016].

[7] D. Chris, "Android SDK suffers from buffer overflow and lack of hardening," 5 February 2014. [Online]. Available: http://www.zdnet.com/article/android-sdk-suffers-from-buffer-overflow-and-lack-of-hardening/. [Accessed 31 October 2016].

[8] V. Kumar, P. Vinay and P. S. Dr Om, "Evaluation of Buffer Overflow and NIDPS," *Internation Journal on Computer Science and Emerging Trends (IJCSET)*, vol. 1, no. 1, pp. 1-8, 2012.

[9] T. D. o. U. C. U. [12] E. J. Schwartz and login: Magazine, "The Danger of Unrandomized Code," *USENIX; login*, vol. 36, no. 6, p. 97, 2011.

[10] D. Gordon, "Address Space Layout Randomization," *Paper Presented at Seminar*, 2009.

[11] H. Shacham, P. Matthew, P. Ben, G. Eu-Jin, M. Nagendra and B. Dan, "On the Effectiveness of Address-space Randomization," in *In Proceedings of the 11th ACM Conference on Computer and Communications Security*, 2004.

[12] S. Bhatkar, D. C. DuVarney and S. Ron, "Address Obfuscation: an Efficient Approach to Combat a Broad Range of Memory Error Exploits," in *In Proceedings of the 12th Conference of USENIX Security Sysposium*, 2003.

[13] E. J. Schwartz, A. Thanassis and B. David, "Q: Exploit Hardening Made Easy," *Proceedings of the USENIX Security Symposium*, pp. 25-41, 2011.

[14] E. J. Schwartz, A. Thanassis and B. David, "Surgically Returning to Randomized lib(c)," *Proceedings of the Annual Computer Security Applications Conference*, p. 60–69, 2009.

[15] Z. Lin, "Recent ROP Defenses: ILR and IPR," *CS 6V81-003: Language-based Security*, 2012.

[16] J. Reavis, "Can a Canary Solve the Buffer Overflow Problem? Network World on Security," 13 August 1999. [Online]. Available: http://www.networkworld.com/newsletters/sec/0830sec1.html. [Accessed 28 July 2013].

[17] R. Chinchani and V. D. B. Eric, "A Fast Static Analysis Approach to Detect Exploit Code Inside Network Flows," *in Proceedings of RAID*, pp. 284-308, 2005.

[18] S. Kalaimagal, M. A. Mukunthan and V. Vijayaraja, "Buffer Overflow Attack Blocker using SIGFREE Concept," *International Journal of Computer Technology and Applications*, vol. 2, no. 2, pp. 365-373, 2011.

[19] X. Wang, P. Chi-Chun, L. Peng and Z. Sencun, "SigFree: A Signature-Free Buffer Overflow Attack Blocker," *IEEE Transactions on Dependable and Secure Computing*, vol. 7, no. 1, pp. 65-79, 2010.

[20] U. A. Sandhu, H. Sajjad, N. Salman and U. A. Obaid, "A Study of the Novel Approaches Used in Intrusion Detection and Prevention Systems," *International Journal of Information and Education Technology*, vol. 1, no. 5, pp. 1-6, 2011.

[21] M. E. W. a. H. J. Mattord, "Security Technology: Intrusion Detection, Access Control, and Other Security Tools," *Principles of Information Security, 3rd ed. , Cengage Learning EMEA*, 2009.

[22] V. Kumar and S. Om Prakash, "Signature Based Intrusion Detection System Using Snort," *International Journal of Computer Applications & Information Technolog*, vol. 1, no. 3, pp. 35-41, 2012.

[23] S. Andersson, C. Andrew and M. George, "Network Based Buffer Overflow Detection by Exploit Code Analysis," *QUT Digital Repositor*, pp. 39-53, 2004.

[24] N. Gaur, "Snort: Planning IDS for Your Enterprise," 11 July 2001. [Online]. Available: http://www.linuxjournal.com/article/4668.

[25] J. Newsome, K. Brad and S. Dawn, "Polygraph: Automatically Generating Signatures for Polymorphic Worms," *In 2005 IEEE Symposium on Security and Privacy (S&P'05)*, pp. 226-241, 2005.

[26] D. J. Day and Z. Zheng-Xu, "Protecting Against Address Space Layout Randomization (ASLR) Compromises and Return-to-Libc Attacks Using Network Intrusion Detection Systems," *International Journal of Automation and Computing*, vol. 8, no. 4, pp. 472-483, 2011.

[27] C. Cowan, F. Wagle, P. Calton, B. Steve and W. Jonathan, "Buffer overflows: Attacks and defenses for the vulnerability of the decade," *In DARPA Information Survivability Conference and Exposition, 2000. DISCEX'00. Proceedings, IEEE*, vol. 2, pp. 119-129, 2000.

[28] K. Hyang-Ah, J. Newsome and K. Brad, "Autograph: Toward Automated, Distributed Worm Signature Detection," *In Proceedings of the 13th Usenix Security Symposium*, vol. 286, pp. 271-286, 2004.

[29] S. Sumeet, E. Cristian, V. George and S. Stefan, "The EarlyBird System for Real-time Detection of Unknown Worms," UCSD, San Diego, 2003.

[30] G. Vigna, R. William and B. Davide, "Testing Network-based Intrusion Detection Signstures Using Mutant Exploits," in *In Proceedings of the 11th ACM Conference on Computer and Communications Security*, 2004.

[31] G. Nalneesh, "Snort: Planning IDS for Your Enterprise," 11 July 2011. [Online]. Available: http://www.linuxjournal.com/article/4668. [Accessed 30 October 2016].

[32] Z. Liang and R. Sekar, "Fast and Automated Generation of Attack Signatures: A Basis for Building Self-protecting Servers," *In Proceedings of 12th ACM Conference on Computer and Communications Security*, pp. 213-222, 2005.

[33] C. Cowan, P. Calton, M. Dave, W. Jonathan, B. Peat, B. Steve, G. Aaron, W. Perry, Z. Qian and H. Heather, "StackGuard: Automatic Adaptive Detection and Prevention of Buffer-overflow Attacks," *In proceedings of 7th Conference of USENIX Security Symposium*, vol. 98, pp. 63-78, 1998.

[34] A. S. Usman, H. Sajjad, N. Salman and U. A. Obaid, "A Study of the Novel Approaches Used in Intrusion Detection and Prevention Systems," *International Journal of Information and Education Technology*, vol. 1, no. 5, p. 426, 2011.

Health Information Management (HIM) System as a Positive Protagonist for Genetic Diseases in Nigeria

(A case study of Sickle Cell Disease (SCD) Patients)

Oluwafisayo Babatope Ayoade
Computer Science Department,
University of Ilorin, PMB 1515,
Ilorin, Kwara State, Nigeria
oluwafisayoadeb.oba@gmail.com

Bamidele A. Oluwade
Computer Science Department,
University of Ilorin, PMB 1515,
Ilorin, Kwara State, Nigeria
deleoluwade@yahoo.com

Taye Oladele Aro
Computer Science Department,
University of Ilorin, PMB 1515,
Ilorin, Kwara State, Nigeria
taiwo_aro@yahoo.com

***Abstract*— Sickle Cell Disease (SCD) which comes in diverse ranges of forms of Hemoglobin disorders (such as sickle cell hemoglobin S-HbSS and hemoglobin C-HbSC, sickle cell anemia (SCA), beta and alpha thalassemias, SD, and SE, etc.). SCD is a common genetic disorder in most sub-Saharan African countries with affecting rate of up to three percent of parturitions (births) as witness in various parts of the continents and especially in Nigeria. It is one of the most uncared non-communicable genetic disorder diseases that are characterized by series of anemia and sickling of red blood cells (RBCs) among Nigerians in recent time. The most common form of SC disorder is produced from the homozygosis of beta-globin S gene mutation (HbSS disease). These red blood cells disorders are portrayed with series of chronic, hemolytic anemia, various injury, vaso-occlusive crisis, stroke, and organ dysfunction in some cases. Patients (which includes children and adults) with SCD are familiar with relate traits and indications, problems and complications of the disease which increases with age. There have been reports of patients' addiction to different types of prophylaxis drugs of SCD such as Tramadol, Pentazocine (Fortwin), Diclofenac injections, etc.; which are prevalent in tertiary institutions. The adequacy of proper data and information management support for managing SCD can reduce the rate of high early-mortality which has been on the rise in children, young adults and adults in Nigeria. The addition of health information management (HIM) programs for the SCD among these pairs will create an affordable method to perform an evaluation of SCD mortality among Nigerians with sufficient health information management.**

Keywords*— *Sickle Cell Disease (SCD), Health Information Management, Hemoglobin, Thalassemia

I. INTRODUCTION

Sickle Cell Disease (SCD) disorder is one of the global health challenges with series of implications which ranges from little to psycho-social levels. In Nigeria today, SCD disorder is one of the most prevalent genetic red blood cells (RBC) disorder affecting citizens of the country of diverse age groups resulting to hemolysis and vaso-occlusive crisis. It was reported that Nigeria has the largest population of people living with SCD disorder with about 150,000 parturitions (births) annually. This survey has also shown that people living with SCD have abnormality in their RBC with their hemoglobin referred to as *hemoglobin S* or *sickle hemoglobin* as featured in their RBC. Human RBC is characterized with this hemoglobin which is the protein that carries oxygen throughout the body system. SCD as a genetic RBC disorder is a non-communicable disease as well as a disease that is transferred through gene from parents to their offspring. It is not communicable like other transmittable diseases like Ebola, Zika, Tuberculosis or common ailments like cold or flu. Patients living with SCD undergo various variability of the disease with different signs and impediments (complications) such as several years of acute pain from birth (i.e. vaso-occlusive episodes (VOEs)), long-lasting pain, several multiple organ injury (e.g. kidney failure, heart failure, etc.), short lifespan, bone infections, stroke among others [1]–[4]. The high frequency of sickle cell anemia disease in Nigeria is one of the highest in the world with an approximation of over 100,000 Nigerian children being given birth to each year with the SC disorder [5]. Patients with this disorder writhe more than average with high frequency of illness and premature deaths especially in children and this can be aggravated to severe attacks such as heart attacks, kidney failure severe bone

infections and sudden mortalities in patients that survive the infancy stage of attacks [5].

In a country of a population estimated to be over 160 million people, it has been statistically shown that over 41 million citizens are carriers of the *'S'* gene of which this approximations total is far above the total number of victims of this carrier in other countries altogether in Africa. In spite of this large statistics of citizens suffering from the SCD disorder or carriers of it, the Nigerian populace continue to demonstrate an uncared attitude to SCD patients and not only this, the general public has a negative perceptions and wrong health information management defiance towards it [5].

National Institute of Health (NIH) in [6] on the SCD disorder demonstrated that people living with SCD disorder genetically inherits two abnormal hemoglobin genes (one from each parent). Generally from research, it has been researched that for all traits of SCD, no less than one of the two abnormal genes causes a person's blood to produced hemoglobin S. In cases where a person has two hemoglobin S genes resulting to the *Hemoglobin SS*, the resulting disease is referred to as *Sickle Cell Anemia (SCA)*. This is the greatest collective and frequently severe type of SCD disorder prevalent in Nigeria. Other forms of SCD observed in Nigeria of recent are Hemoglobin SC disease and hemoglobin S thalassemias (hemoglobin Sα thalassemia, hemoglobin $S\beta^0$ thalassemia, $S\beta^+$ thalassemia, hemoglobin SD, and hemoglobin SE.

In advanced and developed countries like US, UK, Canada and European countries where the SCD disorder is not rampart, the mean life expectancy of individuals living with the homozygous sickle hemoglobin mutation (HbSS) is estimated at 40-60 years and which has improved over the last few decades due to proper health information management [1]. This success development witness an overall increase in Paediatric survival of patients with SCD with the introduction of Prophylactic Penicillin Study carried out in 1986. The discovery demonstrated that the prescription and usage of prophylactic penicillin could alleviate life threatening infections in affected children. Hence, from the study, it was also discovered that collective newborn babies screening became a standard practice especially in US in late '90s and in UK in early 2000 which enabled early diagnosis and patient management [1]. In another development to support proper health information management, introduction of pneumococcal conjugate vaccine which significantly contribute to the decrease in SCD mortality in children younger than 10 years of age were also observed as recorded in [1]. Unfortunately, in less and under developed countries like Nigeria, averafely 50% of children younger than 5 years of age die due to misplaced health diagnosis, complication and non-proper health information management of SCD [1], [2], [7].

Subsequently, more than 98% of children with insignificant or trivial traits of SCD in developed countries are growing and living into adulthood, SCD has become a lingering disorder requiring proper comprehensive life-long management especially in Nigeria. The major challenges of this life threatening disorder facing the adult population in the country include struggles managing the transition from paediatric care into adult care due to lack of available healthcare information management and proper healthcare provision for young adults and adults with SCD. Hence, youths and adults living with SCD disorder depends solely on Emergency Departments (EDs), medical doctors (hematologist), inpatients and outpatients treatment sessions for their care and in some cases these are not available. Improvements have been witness of late in some of the country's University Teaching Hospitals with the emergence of Hematology Departments (HD) to oversee the prevalent scourge of SCD disorder. Therefore the purpose of this paper is to awaken the primary health care medical doctors, hematologists, inpatients and outpatients' doctors, home cares, ED medical doctors and the government with the up-to-date understanding and proper health information management of the SCD through Information and Communication Technology (ICT).

II. Synopsis of Sickle Cell Disease

From medical biology point of view, a *cell* often called the *'building block of life'* is the smallest unit of life that can replicate independently. It consists of cytoplasm enclosed within a membrane containing different biomolecules such as proteins and nucleic acids. Various organisms can be classified either as a *unicellular* (single cell) or *multicellular* organisms (including plants, animals and mammals). While the number of cells in plants and animals varies from species to species, human body contains more than 10 trillion different cells. These cells worked together to sustain life with the support of other non-cellular components of the body such as water, macronutrients (e.g. carbohydrates, proteins, lipids), micronutrients (e.g. vitamins, minerals) and electrolytes. Tissues, on the other hand, are collections of human cells that functioned together in order to perform one single task. Multitudes of tissues function cooperatively to produce an organ that completes specific functions in human body. Despite this structural organization, the entire activities within human body continuously depend solely on the cell. In another perspective, it is a complex unit that makes life possible [1], [6].

Organelles, an important human body component, are the structural component that permits human body to maintain life. They are suspended within a gelatinous matrix called the *cytoplasm* which is controlled within the cell membrane of human body. One of the few cells in the human body that lacks almost all organelles is the Red Blood Cells (RBC). The functions of different cells within human body vary basically on either the type or location of the cells in the human body. The organelles function together to preserve the cell alive and allow it to carry out its specific purpose. Sometimes these organelles are highly dedicated and can vary in size, shape and number. Nevertheless, even though these organelles are utmost functional units, they can neither single handedly exist nor operate alone without the presence of the cell as a whole.

Therefore, in human body, tissue cells require constant supply of oxygen to perform very well. As such, the hemoglobin RBC normally takes up oxygen from the lungs and carries it to the tissues within the body. RBCs containing normal hemoglobin are disc-shaped in nature (doughnut-like shape without a hole), which allows the cells to be flexible for easy movement through large and small blood vessels for the purpose of delivering oxygen. On the other hand, the sickled hemoglobin is not like the normal hemoglobin as it develops stiff rods within the RBCs, changing it into a semicircular or sickle shape. The sickle shaped cells developed are not flexible and as such they stick to the vessel walls of the human cell, causing a blockage that slows or stops flow of blood thereby initiating shortage or total obstruction of oxygen to reach nearby tissues [6] as depicted in Figure 1 (*a*) and (*b*).

The absence of adequate oxygen in tissue cells can initiate series of attacks from mild to unexpected severe pain known as *sickle cell crisis* or *pain crisis*. The crisis attacks can transpire unnoticed and unannounced to the patients and this prompts patients with SCD to visit clinics for effective checkups and treatments. Records have also shown it that most children genetically born with SCD disorder observed pain free occurrences amid painful crisis but this is predominant in young adults and adults with long-lasting enduring discomfort. The presence of sickling RBCs and the absence of adequate oxygen delivery into the tissue cells can results into human body organ damage. Complications such as injury to patient's heart, brain, spleen, eyes, liver, joints, bones, marrow, kidneys, lungs, penis, and skin appear in SCD patients over a lifetime [5], [6], [8]

The resulting inflexible sickle shaped cells form due to lack of oxygen to the tissues can't reshaped easily to the normal flexible type; they tend to burst apart (hemolysis). Normal RBCs have a 90-120 days life span while the sickle cells usually live not more than 20 days (precisely between 10-20 days). Human body always make new RBCs in order to replace damaged old cells due to illness or diseases, as such in SCD patients, the body finds it difficult in renewing these damaged cells, in this case, the number of RBCs is frequently lower than normal resulting into what is known as *anemia* (having less energy) [1], [5]–[7], [9], [10]

Records has shown it that SCD is a life-long disease and its severity varies widely from patients to patients and as such having the right health information and managing the disease well will reduce the high rate of mortality in under-developed countries. Rapid improvements has been witness in developed countries such as US, UK, Canada, etc., witnessing a high life expectancy from 14 years to a range of 40-85 years for SCD patients. Presently, one of the appropriate discovered therapy for SCD patients is the Hematopoietic Stem Cell Transplantation (HSCT), but it is likely unfortunate that most of the patients that can benefit from these cure are either too old for the therapy or could not have a willing and good genetic matched donor as a relative to stand in as a donor for the transplant. For a successful HSCT process, a well matched donor is needed and expected to proffer the best chances for a successful HSCT process. For Nigeria, where the rate of donors are very low or even absent, effective management and treatments can help in reducing the symptoms or crisis and prolong the life of the patients. In another perspective, early diagnosis through proper health information and constant medical care can avert or reduce complications thereby contributing an enhanced healthcare and comfort of the SCD patients [1]–[7], [9], [10]

SCD disorder comes from the output of a single-point transformation (mutation – the replacement of glutamic acid with valine in position 6 on the beta globin (β-globin) subunit of hemoglobin) [1], thereby causing deformation or transmutation of hemoglobin known as *sickle hemoglobin (HbS)*. Patients that genetically inherit two copies of these HbS transmuted or mutations are known as *homozygous (HbSS)* possessing the phenotype form of the disease while those who inherit just one copy of the HbS mutation are called *heterozygous carriers (HbAS)* as they do not exhibit the clinical diagnose disease traits called the *sickle cell trait*. In other words, other types of SCD disorder occurs when mutations responsible for anomalous types of hemoglobin either "*C*" or "*E*" or β-thalassemia combines with HbS as a compound heterozygous mutations producing hemoglobin genotypes such as SC, SE, $S\beta^0$, or $S\beta^+$. It has been noted and reported that people living with HbSS and HbSC are common in Nigeria and also have the greatest unbearable forms of SCD disorder [1], [2], [6], [7], [11].

Abnormal hemoglobin known as *hemoglobin S* results into all the complications experience by a SCD patient. The problem with hemoglobin S is caused by minor deformation in the gene which produces the beta globin (Sβ globin) which changes the way the hemoglobin works [7]. Figure 1 shows a pictorial image of both the normal and abnormal RBCs hemoglobin. The figure also depicts the shapes of the normal and the sickled RBCs hemoglobin [6]

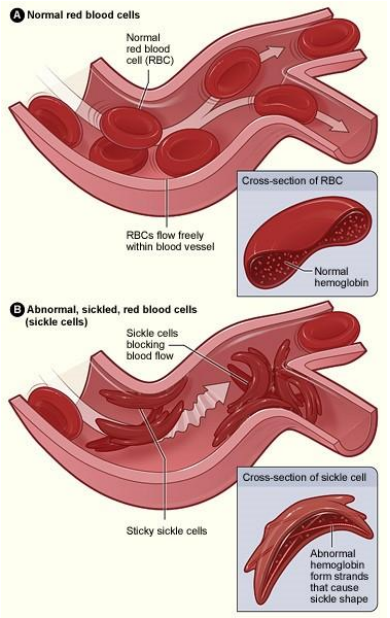


Fig. 8. (a.) Normal RBCs flowing freely showing normal hemoglobin (b.) Abnormal sickled RBCs blocking blood flow in blood vessels showing cross-section of sickled cell with abnormality forming irregular stiff rods [6]

III. GENETICS PATTERN OF SCD

Generally, from medical biology point of view, normal human beings have the following hemoglobin in their bodies. They are:

- hemoglobin *A*: consists of two alpha (α) and two beta (β) chains,
- hemoglobin *A2*: consist of two alpha (α) and two delta (δ) chains and
- hemoglobin *F*: consist of two alpha and two gamma (γ) chains.

From these points, hemoglobin *F* controls the early years of childhood until 6 weeks of age while hemoglobin *A* governs human beings throughout their life time [5], [6].

With this analogy, their sickle-cell conditions in the case of SCD patients possess an autosomal recessive pattern of inheritance from their parents. In this vein, the hemoglobin type a patient inherits from the RBCs depends solely on what type of hemoglobin gene are genetically inherited from their parents. Where one of the parent exhibits sickle cell anemia traits and his or her spouse also exhibits the same sickle cell traits, there is a 50% probability of the person giving birth to a child having a SCD disorder and a 50% chance of having a symptoms or exhibit the sickle cell trait as a carrier. On the other hand, when both parents have a SC trait, the dependents have a 25% chance of SCD, 25% of not carrying the SC alleles but a 50% chance of exhibiting the heterozygous disorder as shown in Figure 2 [1][5][6].

Furthermore, in the case where hemoglobin S gene is inherited from only one parent and a normal hemoglobin gene is inherited from the other parent, there is a tendency that the offspring will have the sickle cell trait as a carrier of the sickle hemoglobin. Children born out of this combination are always healthy; only rarely do people with SC trait have complications similar to those seen in patients with SCD disorder. Nevertheless, people with SC trait are carriers of a defective hemoglobin S gene; hence, there is a probability chance that they can pass it on to their children when they gave birth. On the other hand, in the event that the child's other parent also has sickle cell trait or another abnormal hemoglobin gene like thalassemia, hemoglobin C, hemoglobin D, hemoglobin E, then that child has a chance of having a sickle cell disease disorder.

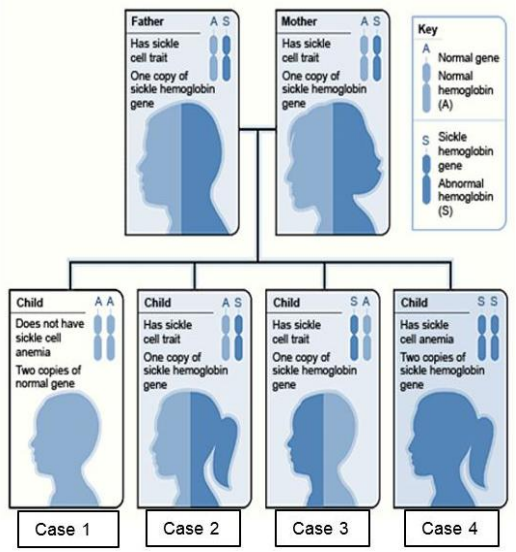


Fig. 9. SCD Inheritance Pattern [6]

SC genes mutation possibly arise naturally in different geographical zones as recorded and proposed by restriction of endonucleic analysis as reported by [12] and [13]. These alternatives country areas such as Benin, Cameroun, Bantu, Saudi Arabia and Senegal has recorded clinical importance because majority of them are connected with higher HbF levels especially in Senegal and Saudi-Asian and they also have a tendency to suffer milder complication of the SCD disorder [12]. However, patients that are heterozygous in nature for HbAS (carriers of sickling hemoglobin), the bonding problems are minor because the normal allele is able to produce over 50% of the hemoglobin. On the contrary, patients exhibiting the homozygous HbSS, the presence of long chain of polymers of HbS distort the shape of the RBC from a smooth doughnut-like shape of the normal RBC to shabby and full points sickled abnormal shapes (as shown in Figure 1(b), rendering them brittle and vulnerable to breaking within the capillaries [5]. SC trait carriers exhibit symptoms of sickling patients only in the event they are stressed up, or their tissue cells are deprived of oxygen (hypoxia), acidosis (increase in acid content in the blood) or severely dehydrated [1][5]. The sickle cell disease crisis follows when the sixth amino acid, called the glutamic acid, is replaced by the valine by changing its structure and function. In another perspective, another name given to sickle cell anemia is E6V. Valine, which is hydrophobic in nature, initiates the hemoglobin to collapse on itself frequently but the arrangement remains the same. Therefore, the numerable amount of collapsing of the hemoglobin on itself causes the RBCs to become sickle in shape [1], [6]. The mutation of single nucleotide of β-globin gene is the gene defects which causes the replacement of the glutamic acid by the valine at position 6 of the hemoglobin chart [1], [3].

The alteration (mutation) caused by the hemoglobin S is known as *HbS* compared to the normal RBCs hemoglobin called the *HbA*. The hydrophobic side chain of the valine

filtrate at position 6 of beta chain is associated with the hydrophobic patch resulting in hemoglobin S molecules to aggregate and form fibroid precipitation in some humans. The allele responsible for sickle cell anemia can be found on the short arm of chromosomes 11 (i.e. 11p15.5) as recorded by [1] and [6]. SCD patients obtains the defective gene from both parents that produces a copy of sickle hemoglobin gene as shown in Figure 2, while patients who receives one defective and one healthy allele remains healthy, but ironically they can pass the disease onto their descendants, who are referred to sickle cell disease (SCD) carriers or heterozygote. However, heterozygotes also are very vulnerable to malaria disease but they experience less severe symptoms of SCD disorder traits [5], [6].

Owing to a very good adaptation benefit of the heterozygote, the SCD disorder is predominant common especially among citizens of the descendants of the lineage of prominent African countries in malaria-troubled areas. Malaria parasites in Nigeria have a complex lifecycle and spend part of its cycle in RBCs. For sickle cell carriers (i.e. *HbAS*), the presence of the malaria parasite in the RBCs of the sickle cell patients causes the defective RBCs hemoglobin to rupture precipitately thereby causing the malaria *Plasmodium parasite* unable to reproduce. Moreover, the process of bonding hemoglobin affects the ability of the malaria parasite to digest the hemoglobin in the first place. Therefore, in areas where malaria is predominantly a common and rampart disease, patients with SCD disorder has a slim chance of survival if they are infected with the SC trait and at the same time attacked with malaria parasite especially the *P. falciparum*. As such, it is important in malaria endemic countries like Nigeria, patients with sickle cell anemia traits (SCA) and particularly children must be protected from malaria by taking necessary and suitable prophylaxis. In less endemic malaria areas like the US, UK, the occurrence of SC anemia among citizens especially African-Americans is lesser (approximately 0.25%) than in Nigeria which is about 4.0% and still declining. With the absence of endemic malaria in high profile zones, sickle cell mutation is honourably not celebrated and tends to reduce in the affected zones like Nigeria by process of natural selection and presently through artificial prenatal genetic screening as observed in US and other developed countries [1], [2], [5], [6], [11]

IV. Physiopathology and Medical Indicators of SCD

RBC elasticity damage is paramount to physiopathology of sickle cell disease (SCD) disorder. Regular or ordinary RBCs are relatively elastic in nature and this gives the tissue cells to buckle or bend in order to penetrate through the human body capillaries as shown in Figure 1. In sickle cell disease patient, inadequate oxygen tension increases the tendency of the RBC sickling thereby the repeated sickling occurrence damage the membrane thereby causing a decrease in tissue cells elasticity. Because of these, the cells find it difficult to reshape to the original or natural shape when the tension in the tissues is finally restored through introduction of oxygen. Therefore, these rigid RBCs finds it difficult to buckle as they penetrate through the narrow passage of the capillaries leading to vessel occlusion (vaso-occlusion in SCD patient) and isochaemia as depicted in Figure 3 [1][6]. The actual SC anemia morbidity is caused by hemolysis (annihilation of RBCs) due to their sickled shape. Even though bone marrows endeavours to compensate by recreating new RBCs, this cannot be compared to the high level of damage that has already been done [6].

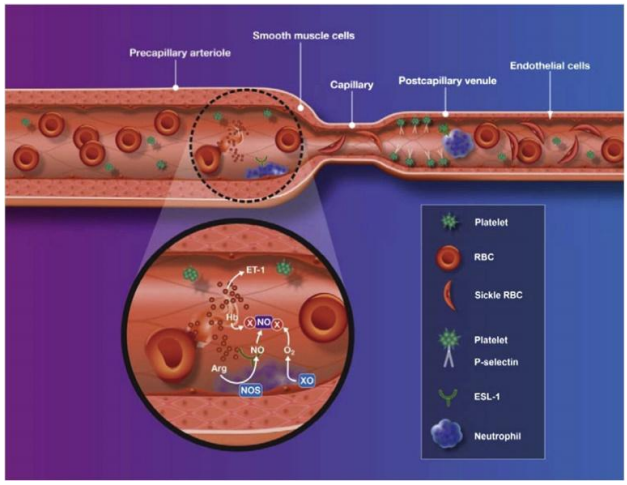


Fig. 10. Physiopathology of SCD vaso-occlusion (Keys: Arg (arginine); ESL-1 (E-selection ligand-1); ET-1 (endothelin-1); Hb (hemoglobin); NO (nitric oxide); NOS (nitric oxide synthase); O2− (superoxide); RBC (red blood cell); XO (xanthine oxidase)) [1], [5]

Medical indicators are very important in SCD management where the sickle cell disorder shows phenotype heterogeneity which results from both genetic and environmental influences. For example, in HbSS, the complete blood counts (CBC) and packed cell volume (PCV) are very vital. Research has shown that CBC reveals hemoglobin levels in the range of 6-8g/dl with high reticulocyte count when the bone marrow tends to recompense for sickled cells destruction by producing more RBCs while PCV on the other hand has its range for male above 18 years of age from 40-52% and female 35-47% in normal hemoglobin. The ranges for the PCV for SCD patients might change alternatively in a decline value with respect to both male and female. However, in other forms of SCD, Hb level are likely to be higher [1], [14]. The abnormality of the hemoglobin can be detected through hemoglobin electrophoresis test (blood test perform to check the different types of hemoglobin present in the blood). Its results are usually ready within several days where various types of hemoglobin move at varying speeds. The sickle cell hemoglobin (HbSS) and hemoglobin C with sickling (HbSC) are major common disorder forms in Nigeria and are detected easily through this test.

One of the medical indicators common to SCD patients in Nigeria is acute or chronic pains which is the trademark of SCD disorder. It can arise out of small vessels blockage or constriction and subsequent tissue infarction (death of tissues due to absence of oxygen), organ damages, or be idiopathic

(occurring without known cause). Vaso-occlusive episodes (VOEs) are painful occurrences that occur from vaso-occlusive crisis with inflammatory and ischemia (restriction of blood supply to tissues) consequences. These VOEs indicators can arise throughout the body of the patient from bones to muscles, mesentery and other organs [1], [14]. Acute sickle cell crisis often precipitate by infections where medical test such as urinalysis (detecting occult urinary tract infection), chest X-ray (occult pneumonia) should be regularly conducted in order to guide against it. In SCD crisis, chronic pain which is poorly understood and managed in Nigeria can be devastating and often times resulted into terrible complications such as leg ulcers, avascular necrosis (death cells due to enzyme degradation), and or neuropathic complications [1].

Acute or chronic pains are not the only medical indicators for SCD disorder. Other defined SCD crisis include aplastic (reduced function of bone marrow), hemolytic crisis (rapid increase in blood cells failure causing PCV level to drop), sequestration (blood pooling in organs), acute chest syndrome (ACS) and stroke (a common recent phenomenon in some patients in Nigeria especially young adults and children) [1], [3], [4], [6], [10], [11], [15][16]. Table 1gives an illustration of major medical indicators of children and adults with SCD and their various expected symptoms.

Genetic counseling is usually rendered to couples planning to get married and have healthy children especially to those that are carriers of the sickle cell disease disorder. For proper, positive and effective counseling, a test is conducted to find out if an unborn child has the disease in the case of pregnant women that are carriers. This is done by either taking blood samples from the fetus or samples of amniotic fluid (pregnant woman's water). However, it has been reported that taking blood samples from the fetus has greater risks than from the amniotic fluid; hence, the latter test is usually used. In addition to genetic or marriage counseling, neonatal screening, which has been successfully demonstrated in developed countries provides not only a method of early detection in pregnant women with SCD but also allows immediate identification in groups of people that carry the SC trait; this can be practice in Nigeria too.

TABLE II. MEDICAL INDICATORS OF CHILDREN AND ADULT SCD (ADAPTED FROM [1])

Children with SCD			**Adult with SCD showing additional indications and problems**
Indications/Symptoms			
	Infants	Children	
	Pain in chest, abdomen, and limbs/joints	Pain (acute or chronic)	Severe joint pain
	Dactylitis	Acute anemia	Chronic leg ulcers
	Anemia	Infections	Retinopathy
	Mild jaundice	Jaundice	Thromboembolic complications
	Enlarged spleen	Poor nutritional status and growth	Neurocognitive impairments
	Fever		Narcotic dependence/tolerance
	Frequent upper respiratory infections	Academic failure	
		Delayed puberty	
Complications			
CNS	Stroke		Recurrent ischemic stroke, hemorrhagic stroke
Eye	Retinal artery occlusion/retinopathy		Progressive retinopathy
Lung	ACS		Recurrent ACS
	Asthma		Pulmonary hypertension
			Chronic lung disease
			Premature coronary artery disease
Heart	Left ventricular hypertrophy		Heart failure
	Cardiomyopathy		Auto-infarction
Spleen	Acute splenic sequestration		Functional asplaenia
	Impaired immunity (e.g. bacterial infection,		
Liver	sepsis)		Hepatic sequestration
			Liver failure secondary to transfusional iron
Kidney			overload
	Hyposthenia		Nephropathy
	Proteinuria		Frequent urinary tract infections
Gall bladder	Renal impairment/failure		
Genitals	Cholelithiasis		Cholelithiasis
Bones/joints	Priapism		Priapism
	Avascular necrosis		Avascular necrosis
Skin	Aplastic crisis		Early loss of bone density
	Chronic ulcers, typically on the ankles		Chronic leg ulcers

Keys: CNS (Central Nervous System), ACS (Acute Chest Syndrome).

Sickle cell disease disorder patients have an increase tendency of being infected and affected by diseases and also they are susceptible to high risk of various life threatening complications such as acute chest pain syndrome, stroke, multi-organ wounds and subsequent end-organ damage, sepsis, pulmonary hypertension, hepatic disease, pulmonary embolism (blockage of an artery in lungs) and cardiomyopathy, etc. The greatest threats of patients living

with sickle cell disease disorders in Nigeria are reduced lifespan, inadequate quality of life due to inadequate health management, high rate of anxiety, poverty, lack of health facilities in rural areas, additions to various prophylaxis, and finally high rate of mortality of patients due to lack of proper healthcare [1], [5]–[7], [10], [14]

V. Signs, Symptoms and Complications of SCD

SCD patients show early signs of the disorder from infants. Every sickle cell disease disorder patients has the disease right from childhood but most children do not have any problem related to the disease until about 5-6 months old while some 8 months. As such, developed countries like US have mandated that all newborn babies receive screening for SCD disorder and parents of SCD patients are notified on time before the child has the symptoms. Some patients show their own sign later in life. In Nigeria, in the past, there has been no stable and adequate programs like the developed countries on managing the SCD disorder early enough especially in the rural areas of the countries but of recent awareness and sensitization has been going on with the advent of hematology departments in various clinics and teaching hospitals in the country. Various foundations and small clubs in the urban areas are trying their best to make sure these process of health information and management is done early enough but the process is not enough or very slow as it pose high risk to the patients and as such increase the rate of mortality [1], [3], [5].

Children or infants exhibiting these sickle cell disease disorder will begin to experience traits of sickle cell problems in their early age days on and symptoms of SCD may include the following among others dactylitis (painful swelling of the hands and legs), anemia fatigue or fussiness, mild jaundice (yellowish color of the skin, nails or eyes), icteris (eyes whitening – occurs when a large number of RBCs hemolyze), enlarged spleen, fever, frequent upper respiratory infection, and so on. These sickle cell disease disorder signs and symptoms varies from one individual to another and can change over time, nevertheless most of them are related to complications of the SCD disorder [5], [6].

In another way, SCD disorder may lead to various acute long-lasting complications and have high rate of mortality [5]. The major complications predominant in Nigeria of recent and most reported ones are as follows which therefore falls in the purpose of this paper which is to awaken the primary health care physicians, hematologist, inpatients and outpatients doctors, home cares guardians to oversee the prevalent scourge of SCD disorder in the country.

A. Vaso-Occlusive Episodes (VOEs)

VOEs crisis is one of the various complications of sickle cell disease disorder caused by sickle-shaped RBCs that hinders the capillaries and restrict body organs blood flow resulting to ischemia, acute pains, necrosis, and sometimes damage to the organs. The incidence, severity and duration of these crises vary significantly from one individual to another. Painful crisis are treated or managed with proper dehydration, correct prescribed analgesics, and in some cases blood transfusions. Appropriate opioid supervision being monitored at regular interval during the crisis until it subsided should be given all through the pain management period. A subgroup of patients' management routine for Non-Steroidal Anti-Inflammatory Drugs (NSAIDS) such as diclofenac, naproxen, Pentazocine (fortwin) can be administered during minor crisis. The later of the NSAIDS (Pentazocine) has been reported of recent to be adversely abused by young adults as it has the same symptoms and components as other narcotic drugs especially for relief in severe pains [5], [6]. VOEs crisis can strike almost anywhere in the body and in more than one point or place of the human body at a time. But often various places in human body that has been recorded where the pain occurs are the lower back, legs, arms, abdomen, chest, and of recent in the brain. These crises can be initiated by illness e.g. fever (prominent), weather/temperature changes, stress, dehydration (not drinking enough water), or at high altitudes. However, in most cases, the patients does not know what triggers or causes the crisis [5], [6], [11], [15], [17]. Young adults and adults living with SCD disorder suffer various complications from chronic pain and it has been shown to be a tough time for individuals to describe these experiences but it is usually different from ordinary pain that occurs from other damaged organ. Some of the VOEs and long-lasting pain if not well managed can lead to high mortality.

B. Severe Anemia

Individual suffering from SCD disorder has different level of crisis. The level of anemia crisis frequency has ranges from mild to moderate levels. In some cases they can develop severe anemia that are life threatening. Severe anemia in children can be caused by splenic sequestration of aplastic crisis.

1) Spleenic Sequestration Crisis

The spleen is one of the human body organs positioned at the upper side of the belly, and its functions are as follows: to filters germs in the blood, breaks up blood cells and develop or recreate new white blood cell. Splenic sequestration crisis occurs in children living with SCD disorder when the RBCs get stuck in the spleen thereby making a quick enlargement which brings about lesser circulation o blood to other cells and this result into anemia crisis. Austere pain may also occur in the left side of the belly to an enlarged big spleen which may cause palpation or wrong feelings by the patient [5].

2) Aplastic Crisis

This is also known as the fifth (5^{th}) disease. It is a mild rash illness caused by Parvovirus B19. It is also referred to as *erythema infectiosum* or *slapped cheek syndrome* because it is the fifth in the list of historical classifications of common skin rash illness in children. Aplastic crisis is common in children than adults. Parvovirus B19 is a very common children infection but in sickle cell disease patients; if fever, itching, running nose, headaches and swollen joints which are its numerous complications are not taken care of properly, it can cause delay or total halt

of bone marrow development of new RBCs for a long period of time which results to severe anemia in children.

Severe anemia is not common to adults living with sickle cell disease disorder but they often experience other causes possibly related to severe anemic conditions such as breath shortage, exhibiting tiredness, dizziness or pale skin.

C. Infections

Human spleen is responsible for protecting the body from certain germs. In human's early life, sickle cells can damage, weaken or destroy the guiding function of the spleen. Patients with sickle cell disease disorder experiences or have damaged spleen at high risk of serious bacterial infections that can be life-threatening. These bacterial infections include but not limited to the following: pneumococcal disease (an infection caused by the Streptococcus pneumonia), hemophilus influenza (bacterial infection common to children under age 5, can be treated by *Hemophillus influenza type b –HiB* vaccine against meningitis), meningococcus (caused by *bacterium Neisseria meningitides* which carries high mortality rate in children if not treated), salmonella (a food poisoning bacteria), staphylococcus (common but usually found on skin, nose, hair and throats of even healthy people), chlamydia (common sexually transmitted disease (STD) that can be easily cured if left untreated but can cause difficulty in women getting pregnant), mycoplasma pneumonia (a form of bacteria that causes infection of the respiratory system), etc. These bacterial infections can cause serious infections from blood infection (septicemia) to lung infection (pneumonia), infection of the covering of the brain and spinal cord (meningitis), bone infection (osteomyelitis). The latter is very common among Nigerians suffering from SCD disorder which has resulted to bone damages or severe bone ulcers leading to amputations [5].

D. Acute Chest Syndrome (ACS)

Sickling of the red blood cells or blood vessels can cause an individual suffering from sickle cell disease disorder to have shortage of oxygen to their lungs. When these occur, the tissues around the lungs get damaged and cannot exchange oxygen properly; and this is known as *Acute Chest Syndrome (ACS)*; where at least one segment of the lungs is damage. ACS frequently exhibits medical symptoms similar to pneumonia. In developed countries like US and UK, ACS has the greatest mortality rate in sickle cell disease patients after 2 years of age, and this is one of the high cause of paediatric intense care unit admissions in developed counters, and recently it was rated the second most collective cause of hospital admission after VOE [1], [5]. ACS is one of the most serious complications of SCD disorder as witnessed in developed countries and should be treated seriously with proper management and treatment in all the hematology departments of Nigerian hospitals. Patients with SCD disorder usually develop these conditions few days after a painful crisis starts. Lung infection may follow these conditions in some cases. There are various signs and symptoms associated with ACS which include painful and sharp chest pain, fever, breath shortage, cough, and in some cases rapid breathing.

E. Stroke

In Nigeria, this complication has been reported of late to be common in SCD patients. This complication is dangerous as it does not give any signs before attacking the SCD patient. Recent records of high morbidity and mortality has been reported with stroke [5], [6]. In some quarters, this has cause emotional breakdown to victims of SCD and their families and oftentimes the loss of life. In HbSS and HbSC patients in Nigeria, the occurrence of overt stroke is about 11% in ages less than 20 years and up to 30% are more frequent in silent cerebral infarct patients [5], [6]. A silent infarct (SI) is known as laceration on magnetic resonance imaging (MRI) consistent with unknown infarction but without focal neurologic deficit lasting longer than 24 hours. Silent infarcts are associated with cerebral weakness, decrease in intellectual abilities, low academic accomplishment, and high risk rate for subsequent infarction (death tissue due to lack of oxygen) [1], [5], [6].

Other complications of SCD disorder are depicted in Figure 4 which includes amongst others brain complications, eye problems, heart disease, pulmonary hypertensions, kidney problems, priapism (common in male - prolonged or painful erection), gallstones, liver complications, bone ulcers especially in legs (osteomyelitis – common in Nigeria), joint complications, hip joints crushing, delayed growth during puberty periods, pregnancy infections, and mental health (frustrations, depression and isolations).

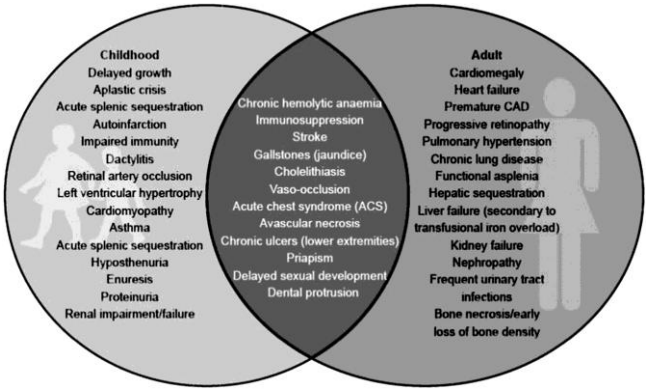


Fig. 11. Sicle Cell disease disorder Complications from Childhood to Adulthood [1] (Keys: coronary artery disease (CAD)

VI. Related Works

Several research works have addressed the study on SCD disorder especially in developed countries and few developing countries particularly in Africa. The most relevant and close related work to this paper is that of [16] where a thorough research review was carried out on management of SCD disorder from childhood to adulthood in the US in 2013. Their work was able to demonstrate that managing SCD disorder and having proper health information on the disease is paramount. They were able to explicate several gaps in

managing the transitions of SCD. Recommendations were also given in their paper review on the next line of action on SCD disorder focusing on discovering cost-effective preventive and comprehensive care measures and programs for adults living with SCD disorder.

Another closely related work is that of [3] where it was stated clearly the process of improving SCD transitions care from childhood to adulthood through health information technology. The survey was carried out also in the US during 2012-2014 through an experimental scan and private interviews with focus groups created to report the success of the development of health information technology (HIT) enabled tool to be used for SCD patients during care transitions. Although, the environmental scan was not a success during the period of the survey as it does not reveal an existing suitable transition tool for the SCD patients, but recorded a minor success where patients, parents, providers, and IT experts saw the potentials, possibilities and applications of creating a tool to meet emergency departments (EDs) health information needs to improve care transitions of these SCD patients.

Other works such as [7], [10] and [15] focused on mortality rate in SCD patients, recent treatment guidelines for managing SCD disorder and morbidity in SCD patients respectively. Thorough studies were carried out on SCD disorder in these papers. [12] was another related study comparing better hematological indices and lower consultation rate of Cameroonian patients' suffering from SCD disorder in a co-heritance between α-thalassemia and sickle cell anemia disease; and their results were positive.

Additional related work that was done in Africa is that of [2] where the focus is on children. Their survey and review was done on available African data on mortality rate in children associated with the most common form of SCD disorder in Africa which is caused by the homozygosis of the beta globin (β-globin) gene S mutation commonly known as *SS* disease. They were able to justify the wrong impression that the SCD condition is associated with the high rate of child mortality in Africa. They carried out their work, in the absence of low contemporary and reliable data, on two studies namely the cross-sectional population survey and cohort studies. They reported that early-life mortality rate was about 50-90% in children with SS disease in Africa and as such involving child existenece procedures and programmes could be of advantage in improving the health information management with respect to death rate among children living with SCD. They also supported the presence of government involvement in creating health investments for child's screening and adequate prophylaxis during their crisis. In same line of mind with this paper, their work also supported adequate and timely data to be available to determine the child mortality in Africa with respect to these interventions and their cost implications. In their conclusion, they recommended blood test samples of already collected specimens of children suffering from SCD disorder covering different age groups with respect to other infections such as HIV, malaria, malnutrition to determine their effect on SCD patients' mortality in African children.

Another work was that of [17] trying to make a comparison between survival of adult patients with SCD in the wealthy environment to the underprivileged zones. The survey was comprehensively carried out in King's College Hospital, London, UK. Their survey also considered on cohort groups subjecting the patients to Hydroxyurea therapy and some to blood transfusion. Their report shows that there are adequate and positive improvements and chances of survival for adult patients living with SCD. Their work was able to calculate mean numbers of survival of patients living with hemoglobin SS disease (HbSS) and other forms of SCD disorder (i.e. HbSC, HBSβ^+, HBSβ^0, and α-globin genotypes) in a wealthy environment in a space of 10 years (from 2004-2013) and it was confirmed that positive result were discovered which is an indication that adequate and proper health information management of SCD disorder can improve life expectancy of victims of the disorder even among adults. Their conclusion was a good landmark for adults living with SCD disorder which shows a life expectancy of 80 years for men and 84 years for women. This survey type is needed in Nigeria in order to get updated information on morbidity and mortality in SCD among adult patients.

Other related works are that of [5]–[11], [13], [14], [19], and [20] showing various epidemiology studies of SCD disorder. The various works were carried out at national and global levels and were able to demonstrate that there is no adequate or timely data estimates of population affected by either the heterozygote (AS) or the homozygous (SS) neonates. Some results were shown demonstrating evidence based estimates from different angles and scales using uncertainty measures.

VII. SCD Management in Nigeria

Comprehensive health information management of SCD patients through Information and Communication Technology (ICT) in Nigeria is of great importance in this dispensation. Various management programs have been in place in the developed countries which has proven that ICT involvement in SCD disorder management have positive impacts on the SCD patients by reducing the complications as well as total annihilation of morbidity and mortality rate. Acute and long-lasting pains are common complications of sickle cell disease patients in developed world and now in Nigeria require individuals living with SCD to seek treatment and necessary information on management of the disorder. Nevertheless, most of these long-lasting pains are due to vaso-occlusion. According to [1] different diagnosis of SCD disorder using systematic approach for its treatment and management of SCD-related pain were discussed. For individuals above 18 years of age in Nigeria, there is scarcity of dedicated resources available especially in the rural areas where acute VOE needs urgent and aggressive treatment [5]. Some customary treatment such as opioids, non-steroidal anti-inflammatory drugs (NSAIDS) such as Pentazocine, diclofenac, etc., and proper hydration can be adequately employed to avert these complications. Table 2 shows a breakdown of both primary

and secondary precautionary measures in young adults and adults with SCD disorder [1].

These programs illustrated in Table 2 can also be implemented in Nigeria and having the right health information management through ICT is supreme. The following are some of the tested and proven management programs for SCD disorder [1][5][6] [20]. They are:

- Folic Acid and Penicillin: Daily intake of folic acid is recommended for every child above 5 years old. In addition, in infants living with SCD disorder, daily penicillin intake, due to immature system of their body makes the system prone to early childhood illness, is also recommended.
- Inhibition of Malaria disease: In Nigeria, malaria attack is one of the prevalent causes of sickle cell anemia crisis in individuals with SCD but this does apply to those with sickle cell traits (HbAS). Therefore, it has been recommended for patients living with SCD, especially in Nigeria, to receive a chemoprophylaxis anti-malarial therapy through their lifetime.
- Vaso-occlusive crisis: This is common in patients with SCD disorder; however the rate of frequency, severity and durations varies from one individual to another. These VOEs are treated with proper pain reducing medications which require pain management with opioids administration at regular intervals during VOEs crisis until the pain subsided. Others can be observed under NSAIDs treatment (e.g. diclofenac, Pentazocine, etc.). In emergency departments (EDs) cases or hospital admissions, intravenous opioids such as patients-controlled-analgesia (PCA) are recommended to be used. In other cases, Diphenhydramine is also recommended to prevent itching associated with the use of some opioids (e.g. the quinine vaccine family for malarial parasite treatment.
- ACS: As discussed earlier, ACS is much more similar to VOEs crisis and as such must be taken seriously and well managed. In such cases antibiotics (e.g. quinolone or macrolide) should be added to supplement for hypoxia (lack of oxygen) [16], [20].
- Hydroxyurea (HU): This is the first approved and established preventive pharmacology treatment for patients living with SCD disorder with persistent VOEs both in infant and adults. In a study carried out in 1995, HU has shown to reduce the frequency rate and severity of VOEs crisis, and also possible increase of survival time in a study carried out in 2003 [20]. The process was done by reactivating the fetal hemoglobin in place of hemoglobin S gene that triggers sickle cell anemia. HU has been previously used as a chemotherapy agent, however, it has been warned that long-term used can be harmful but from research its benefits outweigh the risks [20]. Hence, it is reasonable and risk-taken for HU to be prescribed and managed by primary healthcare doctors through pre-set practice guidelines and hematologist consultation which has been observed and proven in some teaching hospitals in Nigeria of recent.

TABLE III. PRIMARY AND SECONDARY PRECAUTIONARY MEASURES FOR SCD IN INFANTS AND ADULTS (ADAPTED FROM [1])

Precautionary Measures	Management procedure
Immunization – this is highly recommended for HbSS patients with damaged spleen	• Patients with HbSS disease are advised to be immunized yearly against influenza. Intake of pneumococcal vaccine every 5 years, and a meningococcal vaccine series (2-doses primary series controlled 2 months apart for persons under 2 years • Hepatitis A and B vaccines are recommended to patients that have undergone one blood transfusion of the other.
Sickle cell retinopathy screening	• A routine check of this must be performed once a year
Hypertension screening	• This must also be performed at least once annually • Patients should be managed on antihypertensive treatment with the aim of lowering blood pressure to ≤140/90 mm Hg on an average
Proteinuria check (Urine assessment)	• Patients' urine is expected to be checked at intervals of their checkups and in case of tenacious proteinuria, renal specialist is consulted. The use of ACE inhibitors can also be of help.
Pulmonary Hypertension screening	• This is not common with SCD disorder carriers (HbAS asymptomatic) however, it is recommended for sickle cell disease patients (HbSS-symptomatic). Pulmonary hypertension poses a higher mortality risk in vaso-occlusive crisis periods in SCD patients.
Blood transfusion history (10–20 units of blood)	• SCD patients in these categories are at high overload risk • Iron overload measures by ferritin is not always accurate and also increases during long-lasting pains. A level of N1000 ng/mL indicate an overloaded state in SCD patients and patients susceptible to blood transfusion should avoid this while iron chelation should be administered
Ischemic stroke	• The many factors surrounding ischemic stroke in older adults with SCD are not adequately assumed as this has shown similar risk factors in non-SCD patients in this age group • In another perspective, part played by the acute and chronic transfusion in adults has not been adequately defined; therefore,

treatment for stroke in adult patients with SCD is based on past documentation for paediatric patients with SCD. In addition, patients should undergo transfusion exchange for acute stroke through consultation with a hematologist

- Reports have shown that patients living with SCD experiencing stroke may as well have cognitive deficits that cause difficulty in understanding discharge instructions and remembering to keep follow-up appointments. In some cases it leads to death if proper management is not administered

Keys: homozygous sickle hemoglobin mutation (HbSS), andotesin converting enzyme (ACE)

Other management programs include blood transfusion, bone marrow transplant, and proper home management. In support for home management therapy, the following various home management programs for SCD patients are illustrated which include different steps not limited to controlling pain symptoms but also preventing other complications caused by the disease emanating [5]:

- Daily hygiene to children to guide against childhood infection.
- Regular does intake of antibiotic for SCD children up to age 5
- Children immunization on schedule e.g. Hemophilus influenza type B [Hib], hepatitis B, pneumococcal, and influenza vaccine should be given.
- Regular avoidance of dehydration to guide against sickling blood vessels.
- Every SCD disorder patients are encourage to cultivate water intake habit and other fluids before, during, and after exertion and when in the heat.
- Avoid all dehydration factors such as alcohol consumption
- Guide against conditions that reduce the oxygen levels in the blood.
- Guide against long stay in high altitudes e.g. an unpressurized airplane or high mountains
- Avoid cigarette smoking.
- Stress management is highly recommended through adequate sleep and thus reduces fatigue.
- Cold environment and temperature must be avoided e.g. long stay in air-conditioned room or car or long raining periods as this can triggers painful event.
- Proper education on the disease and regular eye examination is recommended
- Develop the practice to identify serious symptoms of the disease. This can be achieved by partnering with your medical doctors and hematologist. Serious warning indicators include fever higher than 101°F (38.33°C); austere cough, difficulty in breathing or shortness of breath; chest pain; severe abdominal (belly) pain; consistent vomiting or persistent diarrhea; sudden increase in the size infants spleen; increased or continuous paleness; lightheadedness; sudden weakness; unexpected numbness or tingling in the hands, feet, fingers, or toes; impulsive development of poor walking balance or coordination.
- Other indicators are confusion; garbled speech or inability to speak; impulsive changes in vision; long-lasting headache; loss of consciousness; persistent erection of the penis (priapism) lasting more than 3 hours or extremely painful; severe pain that can't be relieved with the usual painkilling prescription drugs or other pain-relief methods.
- Consume proper diet and dietary supplements.
- Supplements such as folic acid as prescribed by medical doctors must be observed to aid the development of new red blood cells by the bone marrow.

VIII. Proposed Methodology

The involvement of information and communication technology (ICT) as a positive protagonist in health information management for SCD patients in Nigeria is highly needed for proper data management and health care implementation. This paper work is compiled to address the need to manage and interpret the various amounts of data that have been collated, yet to be collated as well as those generated by different genomic research in order to have a robust health information management through the use of ICT on SCD disorder in Nigeria.

Genes, cells, genetics and genomes are all interrelated terminologies necessary for carrying out this research work in providing a suitable database management for the SCD disorder in Nigeria. Genetics is the study of heredity which originated from Mendel's theory of inheritance of simple traits such as Cowpea colours. One of the major and central abstractions in genetics is the study of gene. Cells, as explained earlier in this paper, is an assessment of chemicals inside a sac bounded by a fatty layer called the *plasma membrane*. Genetics, a material in cells, is contained in a structure called *Chromosomes*. Genome is the hierarchical information that an organism passed to its offspring, represented in each of its cells. The representation is called DNA molecule (*Deoxyribonucleic acid*). DNA is a molecule that carries the genetic instructions used in the growth, development, functioning and reproduction of all living organism and many viruses. Each DNA molecule is a long chain of chemical instructions called *nucleotides* of four different types i.e. fA, C, T, Gg. The totality of all these information is referred to as *GENOME* [21].

In humans, the genomes consist of nucleotides which the molecular biology is given the task to:

- Extract the information contained in the genome of different organisms
- Elucidate the structure of the genome
- Apply this knowledge to the diagnosis and ultimately treatment of genomes disease
- Comparing the genomes of different species and explaining the process and mechanisms of its evolution

Therefore, the proposed methodology to support this paper work is to implement the use of bioinformatics techniques which is the process of using performance computing and mathematics technique to perform a thorough data management, research and analysis of large amount of DNA sequence data in order to find variables that affect the health, disease and drug response with

respect to sickle cell disease disorder in Nigeria. The approach to be implemented is known as *Bioinformatics*.

Bioinformatics relates to biological molecules requiring knowledge in various fields. The threat of over simplifying a very complex issue like SCD disorder, the process of understanding genetic diseases typically profits through three stages

(a.) Recognition of the disease state or syndrome including an assessment of its hereditary character

(b.) Discovering and mapping of the related polymorphisms or mutations.

(c.) Elucidating the biochemical/biophysical mechanisms leading to the disease phenotype

In light of this, this paper would be standing on the methodology of using bioinformatics techniques, a computer-assistant method, to determine the combined effect of ICT on non-communicable disease like sickle cell disease in Nigeria. The proposed method would be implementing SQL – Structures Query Language and SAS (Statical Analysis System) software with MATLAB support to carry out this work on health information management for SCD in Nigeria.

IX. Recommendations

This paper recommends prompt health information management (HIM) programs for SCD in Nigeria with emphasis in the transition stages. Adequate HIM programs through ICT should be employed in hospitals, clinics and rural health centres in Nigeria. Introduction of these HIM programs for appropriate awareness must be inculcate in tertiary institutions against patients' additions to some of the prophylaxis. Better understanding and management of the disorder will greatly reduce the high rate of morbidity and mortality in Nigeria.

X. Conclusion

The SCD disorder is one of the silent killer disease that grossly affects the country, as such, considerable vulnerabilities among patients and accumulating morbidities associate with transitions from childhood to adulthood challenges the proper management of the disease [1]. Management of this type of genetic disease requires proper counseling (adequate health information for individuals and families understand the risks and options), diagnosis (getting the right medical attention) and treatment (averting the morbidity and avoiding mortality of the disease).

In addition, SCD treatments in Nigeria are not adequately and seriously observed unlike other communicable diseases and the use of HU were underutilized compared to other developed countries. Therefore, adult patients with SCD disorder are expected to adequately have access to management services through ICT in various hospitals and clinics to forestall the high rate of morbidity and mortality in Nigeria. The use of health information management system can be the right solution.

Acknowledgment

Medical assistance was provided by the hematologist consultant – Dr. (Mrs.) Ayansanwo A. and ED consultant Dr. Bamgbose of Ekiti State University Teaching Hospital (EKSUTH), Ado Ekiti, Ekiti State, Nigeria, and under proper direction and supervision of the authors. No funding was received from EKSUTH by the authors.

References

[1] J. Kanter and R. Kruse-Jarres, "Management of sickle cell disease from childhood through adulthood.," *Blood Rev.*, vol. 27, no. 6, pp. 279–87, Nov. 2013.

[2] S. D. Grosse, I. Odame, H. K. Atrash, D. D. Amendah, F. B. Piel, and T. N. Williams, "Sickle cell disease in Africa: A neglected cause of early childhood mortality," *Am. J. Prev. Med.*, vol. 41, no. 6 SUPPL.4, pp. S398–S405, 2011.

[3] J. R. Frost, R. K. Cherry, S. O. Oyeku, E. Z. Faro, L. E. Crosby, M. Britto, L. K. Tuchman, I. B. Horn, C. J. Homer, and A. Jain, "Improving Sickle Cell Transitions of Care Through Health Information Technology.," *Am. J. Prev. Med.*, vol. 51, no. 1 Suppl 1, pp. S17-23, Jul. 2016.

[4] M. W. Darlison and B. Modell, "Sickle-cell disorders: limits of descriptive epidemiology.," *Lancet (London, England)*, vol. 381, no. 9861, pp. 98–9, Jan. 2013.

[5] SCAF, "Sickle Cell Disease in Nigeria-Know your Genotype," *Sick. Cell Anemia Found.*, 2016.

[6] N. I. of H. NIH, "Health Information for the Public - Sickle Cell Disease (SCD)," *Natl. Hear. Lung Blood Inst.*, 2016.

[7] O. S. Platt, D. J. Brambilla, W. F. Rosse, P. F. Milner, O. Castro, M. H. Steinberg, and P. P. Klugg, "Mortality in Sickle Cell Disease: Life Expectancy and Risk Factors for Early Death," *N. Engl. J. Med.*, vol. 330, no. 6 June, pp. 1639–1644, 2012.

[8] C. Genet, "Health Research Alliance," vol. 84, no. 2, pp. 1–16, 2014.

[9] L. A. Smith, "Sickle Cell Disease: A Question of Equity and Quality," *Pediatrics*, vol. 117, no. 5, pp. 1763–1770, 2006.

[10] O. S. Platt, B. D. Thorington, D. J. Brambilla, P. F. Milner, W. F. Rosse, E. Vichinsky, and T. R. Kinney, "Pain in Sickle Cell Disease," *N. Engl. J. Med.*, vol. 325, no. 1, pp. 11–16, Jul. 1991.

[11] H. Elmariah, M. E. Garrett, L. M. De Castro, J. C. Jonassaint, K. I. Ataga, J. R. Eckman, A. E. Ashley-Koch, and M. J. Telen, "Factors associated with survival in a contemporary adult sickle cell disease cohort," *Am. J. Hematol.*, vol. 89, no. 5, pp. 530–535, 2014.

[12] M. B. Rumaney, V. Josiane, N. Bitoungui, A. A. Vorster, R. Ramesar, A. P. Kengne, J. Ngogang, and A. Wonkam, "The Co-Inheritance of Alpha-Thalassemia and Sickle Cell Anemia Is Associated with Better Hematological Indices and Lower Consultations Rate in Cameroonian Patients and Could Improve Their Survival," vol. 9, no. 6, pp. 1–10, 2014.

[13] F. B. Piel, A. P. Patil, R. E. Howes, O. A. Nyangiri, P. W. Gething, M. Dewi, W. H. Temperley, T. N. Williams, D. J. Weatherall, and S. I. Hay, "Global epidemiology of Sickle haemoglobin in neonates: A contemporary geostatistical model-based map and population estimates," *Lancet*, vol. 381, no. 9861, pp. 142–151, 2013.

[14] C. J. Homer and S. O. Oyeku, "Sickle Cell Disease: A Roadmap for Getting to Excellence Everywhere.," *Am. J. Prev. Med.*, vol. 51, no. 1 Suppl 1, pp. S3-4, Jul. 2016.

[15] P. Adams-Graves and L. B. Jordan, "Recent treatment guidelines

for managing adult patients with sickle cell disease: challenges in access to care, social issues, and adherence.," *Expert Rev. Hematol.*, vol. 4086, no. September, 2016.

[16] J. Kanter and R. Kruse-Jarres, "Management of sickle cell disease from childhood through adulthood," *Blood Rev.*, vol. 27, no. 6, pp. 279–287, 2013.

[17] K. Gardner, A. Douiri, E. Drasar, M. Allman, A. Mwirigi, M. Awogbade, and S. L. Thein, "Letters to Blood To the editor : Survival in adults with sickle cell disease in a high-income setting," vol. 128, no. 10, pp. 3–5, 2016.

[18] Y. Alapan, C. Kim, A. Adhikari, K. E. Gray, E. Gurkan-Cavusoglu, J. A. Little, and U. A. Gurkan, "Sickle cell disease biochip: a functional red blood cell adhesion assay for monitoring sickle cell disease.," *Transl. Res.*, vol. 173, p. 74–91.e8, Jul. 2016.

[19] K. L. Hassell, "Sickle Cell Disease: A Continued Call to Action.," *Am. J. Prev. Med.*, vol. 51, no. 1 Suppl 1, pp. S1-2, Jul. 2016.

[20] D. Lemogoum, L. Van Bortel, B. Najem, A. Dzudie, C. Teutcha, E. Madu, M. Leeman, J. Degaute, and P. Van De Borne, "Sickle Cell Disease," *Hypertension*, pp. 924–929, 2004.

[21] W. David Mount, *Bioinformatics - Sequence and Genome Analysis*. Cold Spring HarbourLaboratory Press, 2012.

Minimization of Test Cases Using T-Way Interaction Strategy for Efficient Software Testing

Ajiboye, Adeleke Raheem
Dept. of Computer Science,
University of Ilorin,
Ilorin, Nigeria.
ajibraheem@live.com

Mejabi, Omenogo Veronica
Dept. of Information & Comm. Science,
University of Ilorin,
Ilorin, Nigeria.
ovmejabi@gmail.com

Salihu, Shakirat Aderonke
Dept. of Computer Science,
University of Ilorin,
Ilorin, Nigeria.
shaksoft@yahoo.com

Abstract **- One of the key tasks in software development is software testing. In order to develop software that works as expected, extensive testing should be carried out to ensure reliability. Ideally, software testers would want to test every possible permutation of the software but in most cases, due to the complexity of the software, exhaustive testing is usually not feasible. This study applies t-way combinatorial strategy for the reduction in the number of test sets to be used for software testing. The software on which the approach is implemented in this study has parameters of uniform values and their interaction is based on pairwise combination. The technique minimizes the number of test cases by testing all pairs of variables. The test suite generated shows a reduction in the number of test cases by 25%. The final test cases generated are free of redundancy and the technique used shows a high degree of parameter interaction, is efficient and saves the overall time required to test software.**

Keywords: **software testing, software quality, test cases, t-way interaction**.

I. INTRODUCTION

Software is a set of instructions used to acquire set of inputss and to manipulate these acquired inputs for the production of the desired output in terms of functions and performance as determined by the user [1]. In recent times, everyone uses software in one way or another for convenience and better output. We use software in our homes, for business transactions, at the hospitals, and in a number of several consumer products such as cars, mobile phones water dispensers and washing machines. This is why it is said that, it is quite hard to imagine life without software [2]. However, sometimes software may not work as expected as a result of errors or defects. Such errors for instance, may cause the software to fail unexpectedly. The cost of such sudden failure may be too much to bear and may even cause death. Proper testing of software is a promising way to avoid risk, especially risk that is related to human life.

In the process of testing software, the software to be tested is executed with a finite set of test cases, and the behaviour of the system for these test cases is evaluated. The purpose of such evaluation is to determine if the performance of the system meets the expectations. One of the benefits of ensuring that software is extensively tested is that, it gives confidence to the developer team and the users against sudden failure. Although, testing do take place at each stage of software development, further test by independent testers at the end of the software development is crucial to achieving reliable software. The act of software testing helps to measure the quality of software especially as regards to the number of defects found, the tests run, and the system covered by the tests.

In software development, several tasks are performed at different stages and testing is vital to each stage. The most important thing that should be considered in software testing is the design and creation of effective test cases [3]. But due to several parameters that one may need to test in a system, it becomes very important to reduce the number of test cases. In the course of doing this reduction, efforts are made to ensure redundancies are eliminated without sacrificing the reliability of the software being tested.

A test case is a description of conditions and expected results that are taken together in order to fully test a requirement [4]. Ideally, testers would prefer to have all aspects of the software to be covered during testing, but in most cases, especially when the software is complex, exhaustive testing though desirable, is simply not feasible.

In view of the complexity of software to be tested, a meta-heuristic search technique for testing is a realistic approach, this is because, testing every part of the software is practically impossible [5]. The only obvious strategy is to achieve enough tests that can be a representative of the whole test.

In order to address this challenge, the use of sampling testing techniques such as boundary value analysis, equivalence partitioning, decision tables and random testing have yielded some acceptable results [6]. Although, in terms of the likelihood of detecting the most errors in software, the use of random selection to generate test cases have narrow chance of achieving an optimal, or close to optimal subset. This is because the segment that is tested may not be a good representative of the whole partition.

The use of t-way interaction techniques used in this study gives much better variable interaction and minimizes the test cases. A number of t-way strategies exist and these strategies adopt either algebraic or computational approaches. The objective of this study is to demonstrate the use of t-way test strategy in order to minimize the test cases required for the testing of software that has uniform values. The rest of the paper is structured as follows. In the next section, some of the related works reported in the literature is reviewed. This is followed by a brief discussion on the rationales for minimizing test cases. Next to this section is the methodology used for the minimization of the test cases and in Section 5, the result generated is discussed, while the study is concluded in Section 6.

II. RELATED WORK

A number of test cases reduction strategies have been reported in the literature, some of them are reviewed here. Since it is generally known that exhaustive testing of software especially the complex ones is impossible, researchers have focused on the possible ways to have the number of test cases minimized. In order to generate a test suite for software testing, study in [7], combines a greedy algorithm with heuristic search. The study reported that, the suites generated dispense one test at a time and has a good coverage of variables.

In an attempt to minimize the number of test sets, the study in [8] proposed a tree generation strategy for pairwise combinatorial software testing. The technique uses a cost calculation approach iteratively and all the leaf nodes were considered. The proposed approach was reported to have been used to generate the test suite until all the combinations were covered. A significant reduction in the number of test cases was also reported.

Also, in order to reduce the size of test suites generated for testing purposes, the study described in [9], applied particle swarm optimization technique which is a kind of meta-heuristic search technique to pairwise testing. The study further proposed two different algorithms for efficient generation of test cases.

Furthermore, the study in [10], proposed along with other techniques, a backtracking algorithm and search heuristics. Both techniques were found useful in the generation of test suites for combinatorial testing. The study reported that the proposed method was suitable and efficient in the reduction of the size of test cases. Another study focuses on using the interaction-based test-suite minimization approach as a way of standardizing combinatorial test design [11].

A number of sampling methods for the minimization of test cases have also been reported in the literature. Examples of the techniques in this category include: equivalence partitioning, boundary value analysis and random testing. Reid [6], compares some of these methods and found boundary value analysis to be the most effective among the techniques that were investigated.

The study reported in [12], relies on the use of sampling methods for the construction of an automated framework for test data generation. Equivalent Partitioning (EP) technique is an example of sampling method. EP technique reduces the number of test sets since the values in an equivalence partition are handled similarly. Testing only one part is assumed to have catered for the entire partition. One of the challenges with the use of sampling technique is the possibility of not achieving the optimum results. This is because the sample data used for testing may not be a good representative of the domain test data.

III. RATIONALES FOR MINIMIZING TEST CASES

A test case can be defined as a set of conditions or variables under which a tester will determine whether a system under test satisfies requirements or works as expected [13]. Due to a combinatorial explosion in software testing, it is necessary to carefully minimize test cases to a manageable size that can still give optimum results. For instance, considering software of 15 parameters, 14 of these parameters have 2 values each and the 15^{th} parameter has 65 values; this may be a set of different colours. This would give a combinatorial explosion that can take several years to test: $2^{14} \times 65 = 16{,}384 \times 65 = 1{,}064{,}960$ combinations.

Since it is practically impossible to test large combinations like this exhaustively, then a minimized test set is inevitable. Some reasons why software may not work as expected include the environmental conditions like the presence of radiation, magnetism, electronic fields or pollution [14]. Further clarifications made in [14] shows that these factors are capable of changing the way hardware and firmware operate. Such sudden change can lead to system failure. In order to avoid system failure, therefore, we should critically look for errors and faults in a system with a view to rectifying them.

One of the implications of software that is not working correctly is the possible potential of causing harm to people. It can make a company to record high loss and can lead to pollution of the environment [14]. For instance, incorrect billing can lead to huge losses to a company, while the release of radiation or poisonous chemicals to the environment can cause serious disaster.

The procedure involved in making high-quality products is very tasking. Although, certain costs may be obvious, like sales, marketing, and employee salaries; other costs, especially indirect ones such as employee training, testing, fixing errors, retesting, and reporting on information, may be less obvious and are often neglected by organizations; and of these, testing appears to be the most bemoaned [15].

Test planning that focuses on creation and design of test cases is perhaps the most pivotal aspect of software testing [2]. The reason for this is that, a test design that is poorly planned is capable of making the bug detection in the system difficult. Test-case design is so important because it is not possible to test all; and it follows that, a test of any program must be necessarily incomplete [3].

There are several procedures that must be followed in the course of designing test cases. Even after selecting the set of test cases necessary to cover the identified risks, there are still other things that should also be considered necessary. Test cases may need to be changed, adapted, or added and in addition, careful attention must be given to ensure that test data can be reused [16]. In this phase, effort is expended in setting up the test data for test cases.

The writing of test cases should encompass the description of the main functionality that needs to be tested and the readiness to assure that the test can be conducted. What make up the test cases basically includes the set of input values, execution preconditions, expected results, and execution post conditions, developed for a particular objective or test condition; for instance, to exercise a particular program path or to verify compliance with a specific requirement [17].

IV. METHODOLOGY

T-way pairwise interaction technique is used in this study for the reduction in the number of test cases. There are a number of methods that can be used under t-way combinatorial techniques. The number of values is one of the determinants of the approach that is most suitable. Since the software under test represented in Figure 1, has uniform values, t- way pairwise appears to be a suitable technique. The number of interactions, t, is taken as 2. Figure 1, is an excerpt from a system and the users are expected to choose a value from each parameter to place an order. The single interface under test has 3 parameters each with 2 values. That is, X = x_1, x_2 ; Y = y_1, y_2 ; Z = z_1, z_2. The values x_1, x_2; y_1, y_2 and z_1, z_2 are referred to as base values and assigned to the parameters as:

Crust = {Classic Hand Tossed (x_1), Crunchy Thing (x_2)};
Flavour = {Vegetarian (y_1), Pepperoni Delight (y_2)};
Toppings = {Pineapples (z_1), Beef (z_2)};

A test suite is constructed with these parameters as represented in Table 1. The mathematical notation of the technique used conforms to the expression in (1):

$$CA\ (N,t,V^P) \qquad (1)$$

where CA is the Covering Array, N is the size of test suite, *t* is the number of interactions. In this instance, *t* = 2.
V is the number of values.
P is the number of parameters.

Therefore, using covering array and taking the interaction, t = 2, the number of test cases that can be generated for exhaustive testing is V^P. Since the number of values for each parameter is 2, the exhaustive testing for the interface represented in Figure 1, becomes 2^3, which equals 8 test cases. The exhaustive combination is illustrated in Table 1.

It should be noted that the possible techniques that can be used under t-way interaction is not limited to uniform strength interaction used in this study. Other known interaction strategies are: cumulative strength, variable strength and input-output based relation. In the course of choosing any of these techniques, the parameters and the distribution of values is usually taken into consideration. The tables 1, 2, 3 and 4 are the general template for an interface of 3 parameters, each with a pair of uniform values. Apart from using x, y, and z, any other characters can also be used to denote the values; digits can also be used instead as far as it brings about clarity.

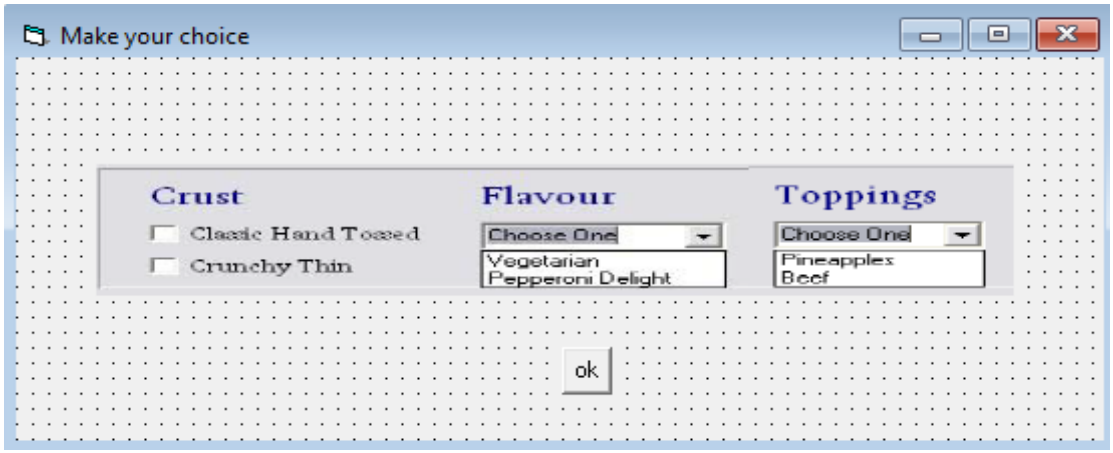


Fig. 1. Interface showing parameters of uniform values

Table 1. Exhaustive 2-way combination

	Input		
	X	Y	Z
Base Values	x_1	y_1	z_1
	x_2	y_2	z_2
Exhaustive Combinations	x_1	y_1	z_1
	x_1	y_1	z_2
	x_1	y_2	z_1
	x_1	y_2	z_2
	x_2	y_1	z_1
	x_2	y_1	z_2
	x_2	y_2	z_1
	x_2	y_2	z_2

The equivalent test set is as follows:
1: { Classic Hand Tossed, Vegetarian, Pineapples}
2: { Classic Hand Tossed, Vegetarian, Beef}
3: { Classic Hand Tossed, Pepperoni Delight, Pineapples}
4: { Classic Hand Tossed, Pepperoni Delight, Beef}
5: { Crunchy Thing, Vegetarian, Pineapples}
6: { Crunchy Thing, Vegetarian, Beef}
7: { Crunchy Thing, Pepperoni Delight, Pineapples}
8: { Crunchy Thing, Pepperoni Delight, Beef}

For an interface of three parameters, for instance XYZ as represented in Figure 1, the possible permutations are: XY, XZ and YZ. Each of these combinations forms the set of test cases. What is unique about 2-way testing technique is that, the pairwise technique minimizes the number of test cases by testing all pairs of variables. Thus, the combination of each pair of variables gives the table shown in Tables 2, 3 and 4.

The number of occurrences in each pair of combination is determined. The focus is to ensure that no variable is uncovered, while those that appear more than once are removed to avoid redundancy as illustrated in Tables 2 and 3. The left over, that is, those that appear only once form the final test suite shown in the result section (see Table 5). Although, there are several software testing tools that can be used to generate test suite, however, this procedure is implemented using Matlab codes.

Table 2. The combination of parameters X and Y

	Input		
	X	Y	Z
Base	x_1	y_1	z_1

Table 3. The combination of parameters X and Z

	Input		
	X	Y	Z
Base Values	x_1	y_1	z_1
	x_2	y_2	z_2
XZ	~~x_1~~	~~y_1~~	~~z_1~~
	x_1	y_1	z_2
	x_2	y_2	z_1
	~~x_2~~	~~y_2~~	~~z_2~~

Table 4. The combination of parameters Y and Z

	Input		
	X	Y	Z
Base Values	x_1	y_1	z_1
	x_2	y_2	z_2
YZ	x_1	y_1	z_1
	x_2	y_1	z_2
	x_1	y_2	z_1
	x_2	y_2	z_2
Values	x_2	y_2	z_2
XY	~~x_1~~	~~y_1~~	~~z_1~~
	~~x_1~~	~~y_2~~	~~z_1~~
	~~x_2~~	~~y_1~~	~~z_2~~
	~~x_2~~	~~y_2~~	~~z_2~~

Table 5. The final minimized test suite generated

	Input		
	X	Y	Z
Base Values	x_1	y_1	z_1
	x_2	y_2	z_2
Pairwise	x_1	y_1	z_1
	x_1	y_1	z_2
	x_1	y_2	z_1

Combinatorial	x_2	y_1	z_1
	x_2	y_1	z_2
	x_2	y_2	z_2

V. RESULT AND DISCUSSION

This section discusses the test suite generated from the pairwise interaction of parameters under test. From the initial exhaustive testing, the study further reduces the number of test cases.

The final result suite is represented in Table 5. The result shows some significant reduction in the number of test cases from an initial test suite of size 8 as earlier shown in Table 1, to 6 test cases. This is equivalent to 25% minimization of the number of test cases.

Those test cases that occurred twice, which could have resulted to duplications have been removed from the permutation shown in Tables 2 and 3. Only the unique test cases are left and harnessed together to form Table 5. The selection of the parameter values shown earlier in Figure 1 can then be made based on these final cases without ignoring or making repetition in the design of the test cases.

The selection of test cases in the software excerpt represented in Figure 1, gives the exhaustive combinations of 3 parameters, each with 2 uniform values. The interaction of the parameters based on the possible permutations and subsequent elimination of redundancies results to unique test sets. The initial test suite is then minimized to the following test cases as listed in numbers 1 - 6:

VI. CONCLUSION

This paper has shown how the t-way interaction technique can be used to minimize the number of test cases in the process of testing software. When a test is carried out on software, such effort can only show that one or more defects exist. Testing is not capable of showing that the software under test is error free. The essence of software testing is, therefore, to greatly reduce risk or sudden failure of the software.

Due to the challenges one is bound to face when the software is to be tested exhaustively, the most feasible approach is to minimize the test sets without sacrificing the reliability of the software. Testing unveils some inconsistencies and bugs that may find its way into the software. In reality, software that consists of large parameters of non-uniform values can take some decades to test, hence, the need to have the test sets minimized.

Although the interface under test represented earlier in Figure 1, has a simple logical structure, it is for illustrative purpose, as it consists of 3 parameters of 2 uniform values only; however, a larger number of parameters of uniform values follow the same procedures shown in this paper. The results from this study show that, the number of test sets generated for the exhaustive combinations of the parameters have been reduced by 25%. Apart from the fact that the approach eliminates some duplication that is capable of swelling up the number of test cases, thereby saving testing time, the technique is found to be efficient in the minimization of test sets.

ACKNOWLEDGMENT

The Authors would like to thank the anonymous reviewers for their useful comments. We also appreciate the support by the department of Computer Science, University of Ilorin, Ilorin.

REFERENCES

[1] B. B. Agarwal, S. P. Tayal, and M. Gupta, *Software Engineering & Testing*: Jones and Bartlett Publishers, 2010.

[2] K. Z. Zamli and B. Y. Alkazemi, *Combinatorial T-way Testing*. Malaysia: UMP, 2015.

[3] G. J. Myers, *The Art of Software Testing*, Second ed.: John Wiley & Sons, 2004.

[4] G. Fournier, *Essential Software Testing A Use-Case Approach*: CRC Tailor & Francis Group, 2009.

[5] W. Afzal, R. Torkar, and R. Feldt, "A systematic review of search-based testing for non-functional system properties," *Information and Software Technology*, 2008.

[6] S. C. Reid, "An empirical analysis of equivalence partitioning, boundary value analysis and random testing," in *Fourth International Software Metrics Symposium*, 1997, pp. 64-73.

[7] R. C. Bryce and C. J. Colbourn, "One-test-at-a-time heuristic search for interaction test suites," in *9th annual conference on Genetic and evolutionary computation*, 2007, pp. 1082-1089.

[8] M. F. Klaib, S. Muthuraman, N. Ahmad, and R. Sidek, "Tree based test case generation and cost calculation strategy for uniform parametric pairwise testing," *Journal of Computer Science*, vol. 6, p. 542, 2010.

[9] X. Chen, Q. Gu, J. Qi, and D. Chen, "Applying particle swarm optimization to pairwise testing," in *34th IEEE Annual Computer Software and Applications Conference*, 2010, pp. 107-116.

[10] J. Yan and J. Zhang, "Backtracking algorithms and search heuristics to generate test suites for combinatorial testing," in *30th Annual International Conference on Computer Software and Applications* 2006, pp. 385-394.

An Adaptive Yoruba-English Word-Search Puzzle

Olaniyi Tawakalt Taiwo
Department of Computer Science
University of Ilorin
Ilorin, Nigeria.
olaniyitaiwo95@yahoo.com

Yetunde Folajimi
Department of Computer Science
University of Ibadan
Ibadan, Nigeria.
yetundeofolajimi@gmail.com

Bamidele A. Oluwade
Department of Computer Science
University of Ilorin
Ilorin, Nigeria.
deleoluwade@yahoo.com

ABSTRACT: Upholding heritage language is the key to upholding next generation. Yoruba Language which is one of the four official languages of Nigeria is spoken by over 25 million people in the world, some of which are in diaspora. There are not enough software tools to help preserve the language, and as much as people like to solve puzzles, there are only few of them in their native language. Hence, the need for a Yoruba-English word-search puzzle. This paper centers on the development of an Adaptive Yoruba-English word-search puzzle for enhancing the learning of Yoruba words and their English equivalence. The Methodology include; Requirements gathering, reviewing existing word-game systems, developing a system models to visualize the architecture of the system and implementing the system. The new system was designed using Hypertext Markup Language, JavaScript, CSS, hypertext preprocessor (Php) and the Application programming Interface used was GameMaker. These languages and tool were chosen because of their easy syntax and features for developing web based game applications. The design and Implementation of this paper lead to the development of a web based Adaptive Yoruba-English Word-search Puzzle, thus making the learning of Yoruba words fun and interesting.

Keywords: Language, Yoruba Language, Puzzle, Word-Search Puzzle, Vocabulary.

I. INTRODUCTION

Language is a systematic means of communicating ideas or feelings by the use of conventionalized signs, sounds, gestures, or marks having understood meanings. (Douglas, 2007). Learning a new language from scratch can be said to be very tricky and interested learners often get bewildered about where to start. Do they have to first learn grammar, conversation, vocabulary, reading or writing? A good way to learn a language, especially vocabulary is by using games, such as a word puzzles.

A **puzzle** is a game, problem, or toy that tests a person's ingenuity or knowledge. In a puzzle, one is required to put pieces together in a logical way, in order to arrive at the correct solution of the puzzle. Puzzles are often devised as a form of entertainment but they can also arise from serious mathematical or logistical problems. In such cases, their solution may be a significant contribution to mathematical research. (Kendall, Parkes, & Spoerer, 2008). Solutions of puzzles often require the recognition of patterns and the creation of a particular kind of order. People with a high level of inductive reasoning aptitude may be better at solving such puzzles than others. (Carter & Russell, 1993) But puzzles based upon inquiry and discovery may be solved more easily by those with good deduction skills. Deductive reasoning improves with practice. (Peterson & Simon, 1998)

Right from early childhood to adulthood, individuals love to play with puzzles. The way puzzles challenge our thinking and exercise our minds is thrilling and exciting. Puzzles are helpful in the growth and development of young people. It was however recognized that there are only few Yoruba learning apps, and that learning Yoruba was restricted to formal learning (in schools and learning institutions) and informal learning (indirectly from Yoruba speaking people) due to the problems stated above, hence, the need arose to develop a Yoruba-English word-search puzzle with functionalities to make Yoruba learning fun and interesting.

This study was carried out to help people familiarize themselves with more Yoruba words and their English equivalence in a fun and interesting way with the aid of puzzles. Players learn more Yoruba words starting from the basics.
The deliverable of this paper is web application which applies the concept of puzzles to the study of Yoruba words and their English equivalence. The client interface generates a set of words randomly from the available words. The age and gender of the player determines interface, puzzle grid size and category of word to be displayed

II. PROBLEM STATEMENT

Upholding the heritage language of a tribe is the key to upholding and giving a true sense of identity and a high level of self-esteem to the next generation.
Yoruba Language which is one of the four official languages of Nigeria is spoken by over 25 million people in the world, some of which are in diaspora. There are not enough software tools to help preserve the language, and as much as children like to solve puzzles, there are only few of them in their native

language. Hence, the need for a Yoruba-English word-search puzzle.

III. ADAPTIVE YORUBA-ENGLISH WORD-SEARCH PUZZLE

A. FUNCTIONAL REQUIREMENTS

The system performs the following functions:

1. The system should be able to generate as many words.
2. The system should be able to capture user's age and sex.
3. The system should be able to scramble alphabets making up words efficiently.
4. The system should be able to determine appropriate user interface for different gender.
5. The system should be able to determine appropriate classes for different age groups.
6. The system should provide a timer to measure the player's speed while solving the puzzle.
7. The system should be able to solve the puzzle incase the player encounters any difficulty.

B. NON-FUNCTIONAL REQUIREMENTS

The proposed system will provide the following non-functional requirements:

1. The system should be easy to understand by all age groups.
2. The system should be fun to use (play).
3. The system should be interactive so that players can have a true sense of participation.
4. The system should be very educative.

IV. DESIGN

The major focus of this paper is to develop a word-search puzzle with a number of operations to be implemented. In this chapter, the development of the design model is strictly based on Unified Modelling Language (UML) diagrams.

5. The system should be able to pick up from where a player left off.

C. PLAYER MODULE

In this module, the various operations can be performed;

1. The player should be able to input his/her details.
   A form should be displayed at the home page of the puzzle to capture player's details.
2. The player should be able to play the game by searching for the English equivalence of the given Yoruba words.
   Different Yoruba words should be displayed, with their English equivalence scrambled in the puzzle grid. The player will then be required to search for the English equivalence of the listed Yoruba words.
3. The player should be able to resize the Puzzle grid.
   Enlarge and Reduce buttons should be included to allow players zoom in and out the puzzle grid.
4. The player should be able to ask the system for help.
   A help button will be included to give instructions as to the operations of the system. Also, solve buttons will be tied to all listed Yoruba words to allow the system help find the correct English equivalence.
5. The player should be able to restart the puzzle.
   A restart button will be added to the game to allow players start the puzzle afresh with newly generated words.

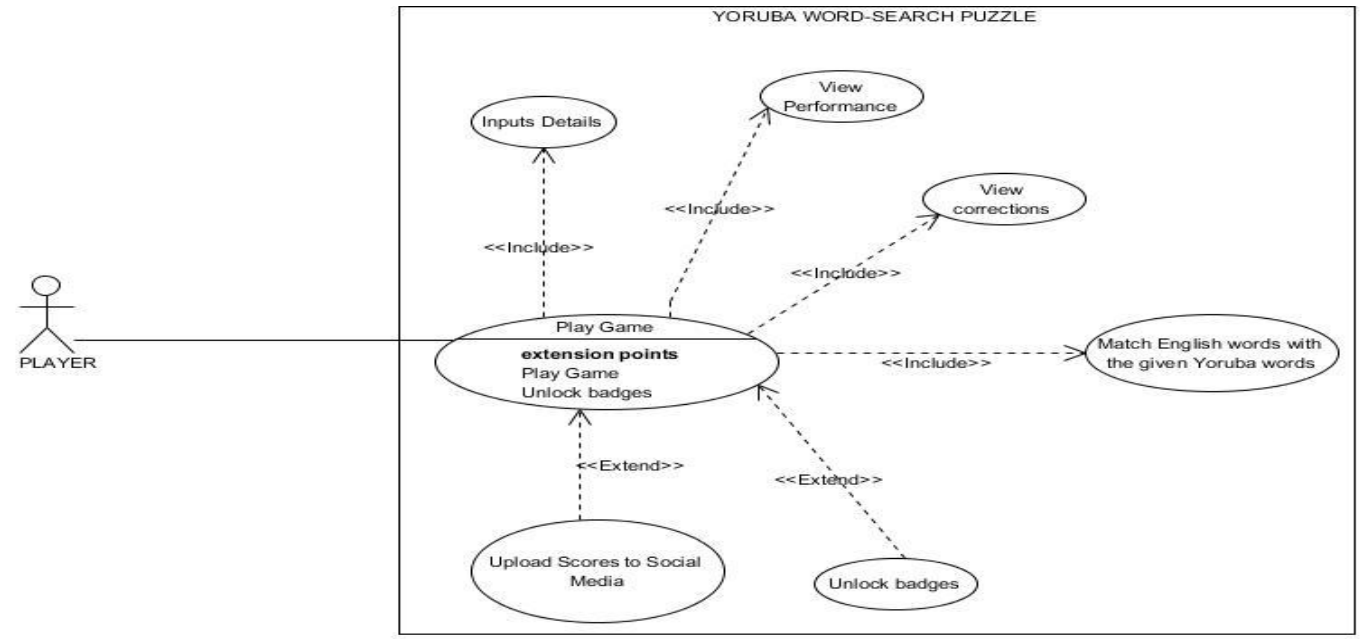


Figure 1: Use Case Model

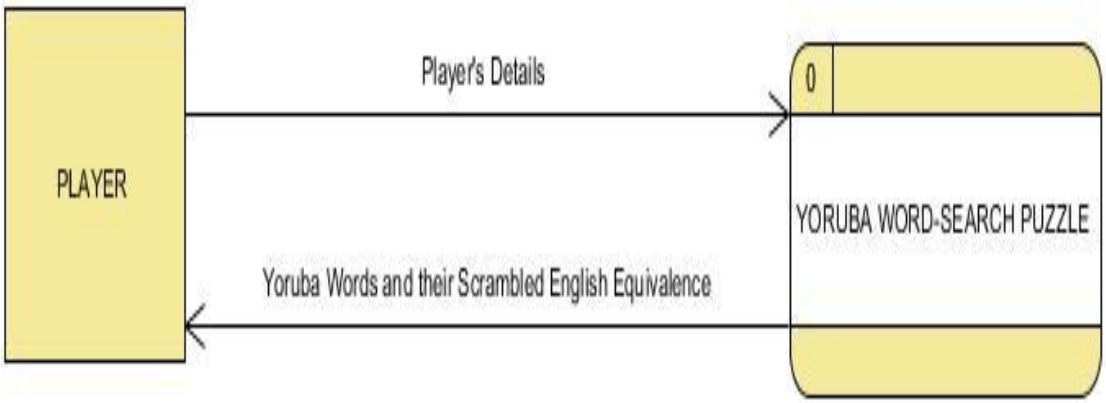


Figure 2: Context Diagram

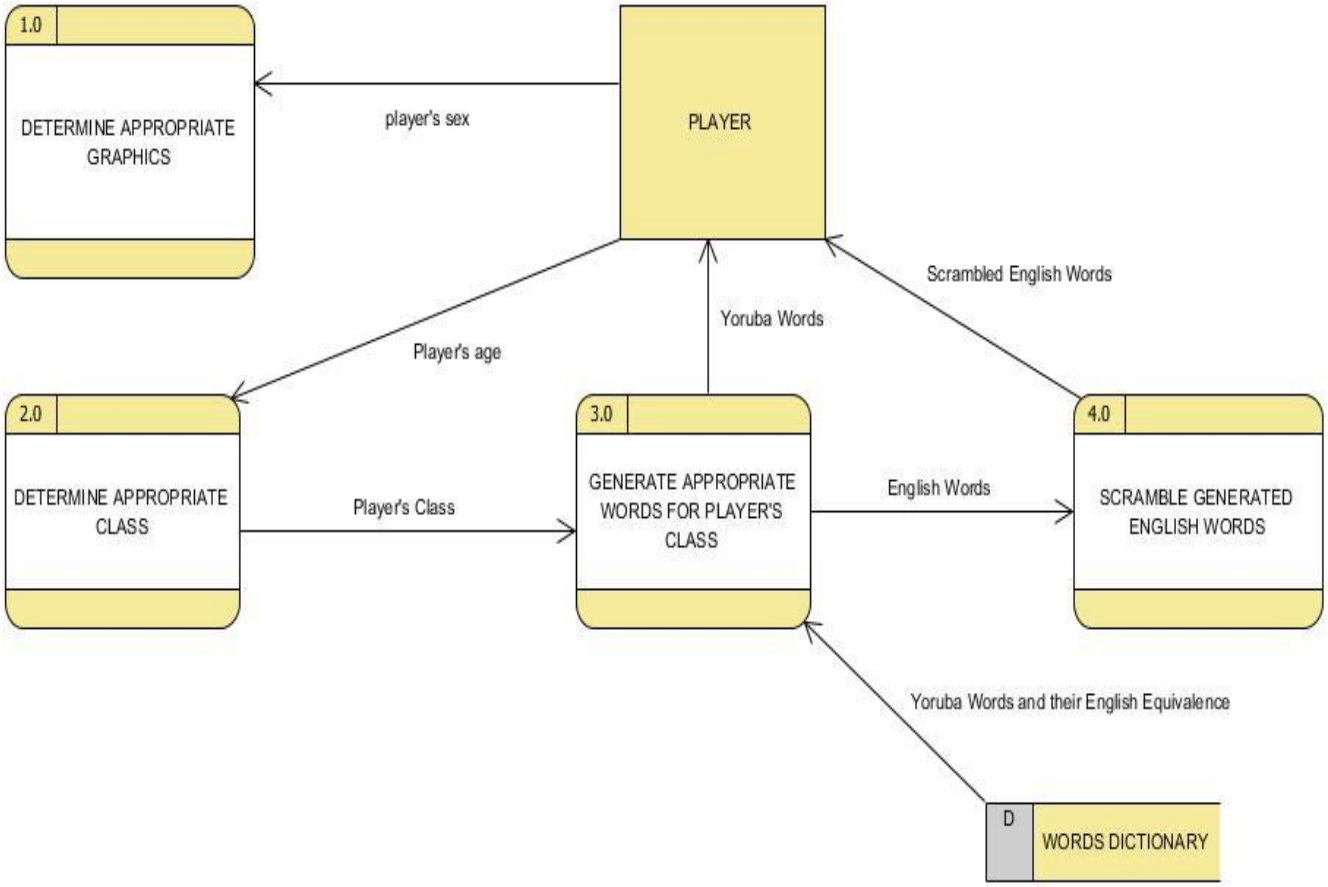


Figure 3: System DFD

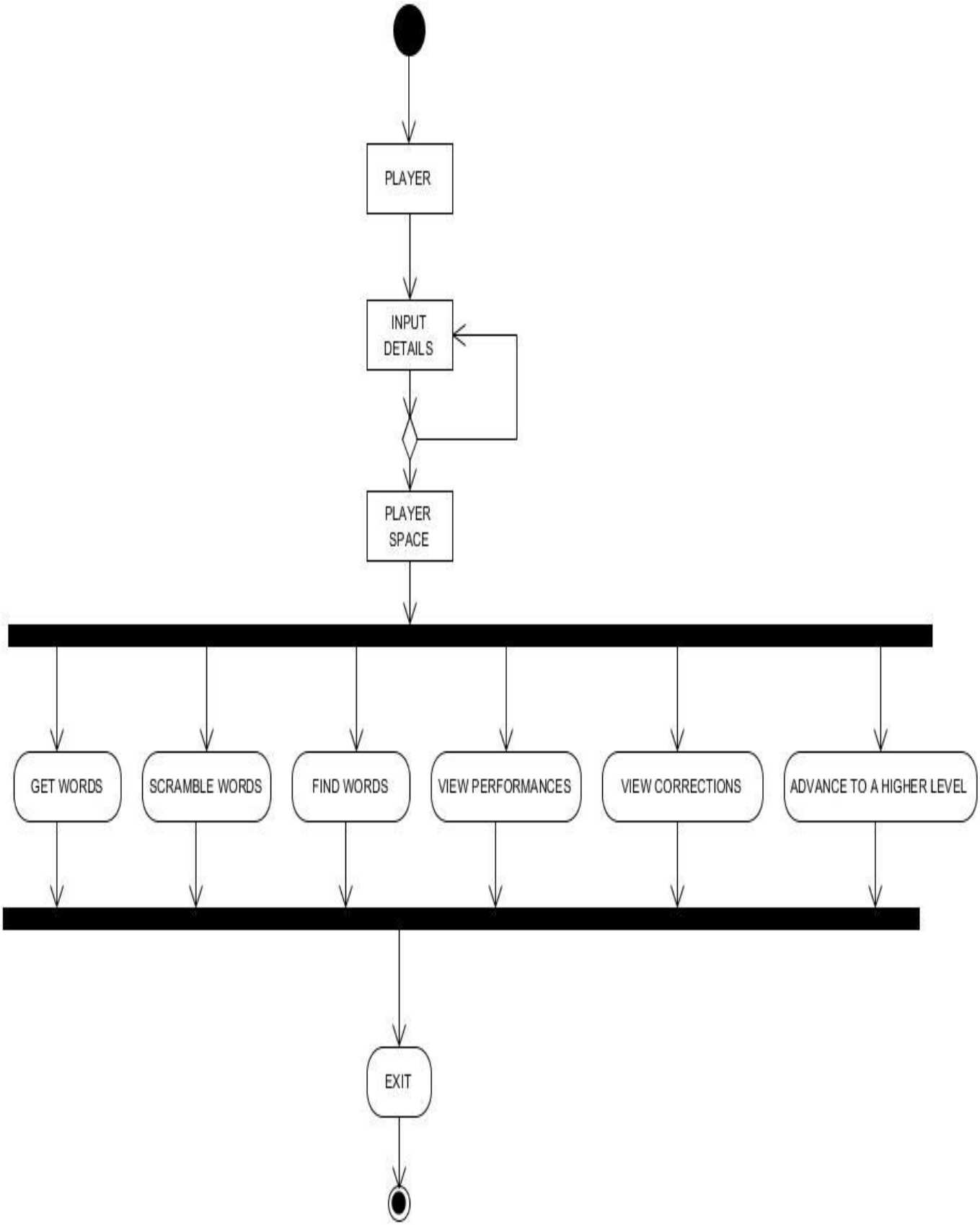

PLAYER
INPUT DETAILS
PLAYER SPACE
GET WORDS
SCRAMBLE WORDS
FIND WORDS
VIEW PERFORMANCES
VIEW CORRECTIONS
ADVANCE TO A HIGHER LEVEL
EXIT

Figure 4: Activity Diagram

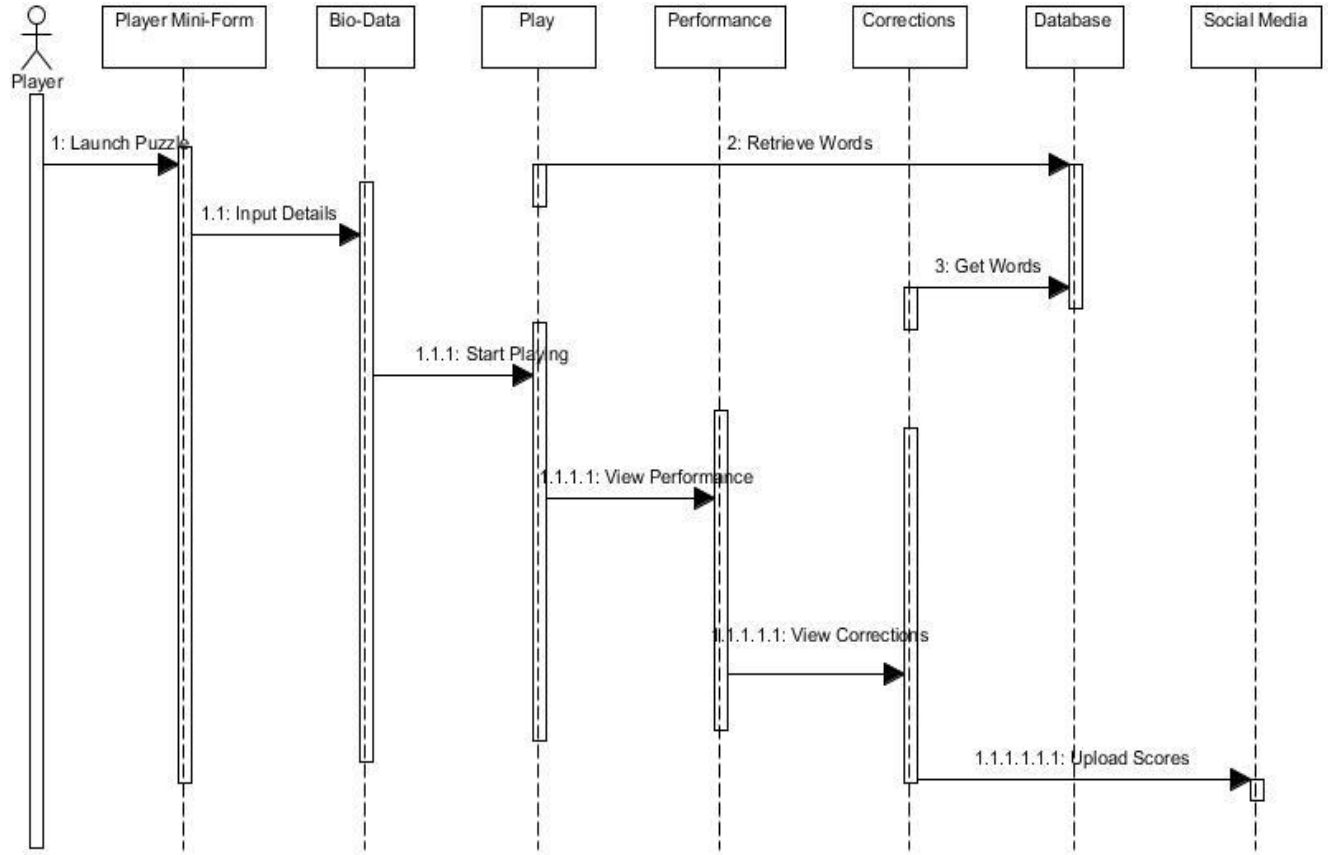


Figure 5: Sequence Diagram

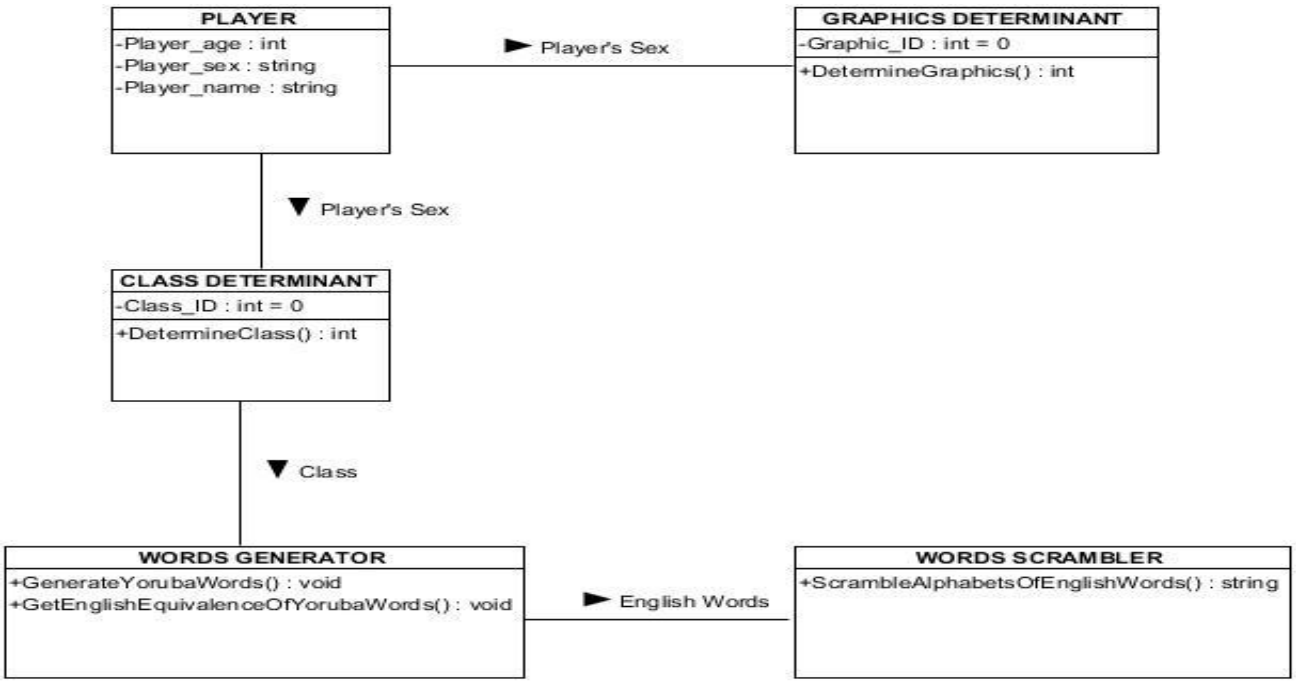


Figure 6: Class Diagram

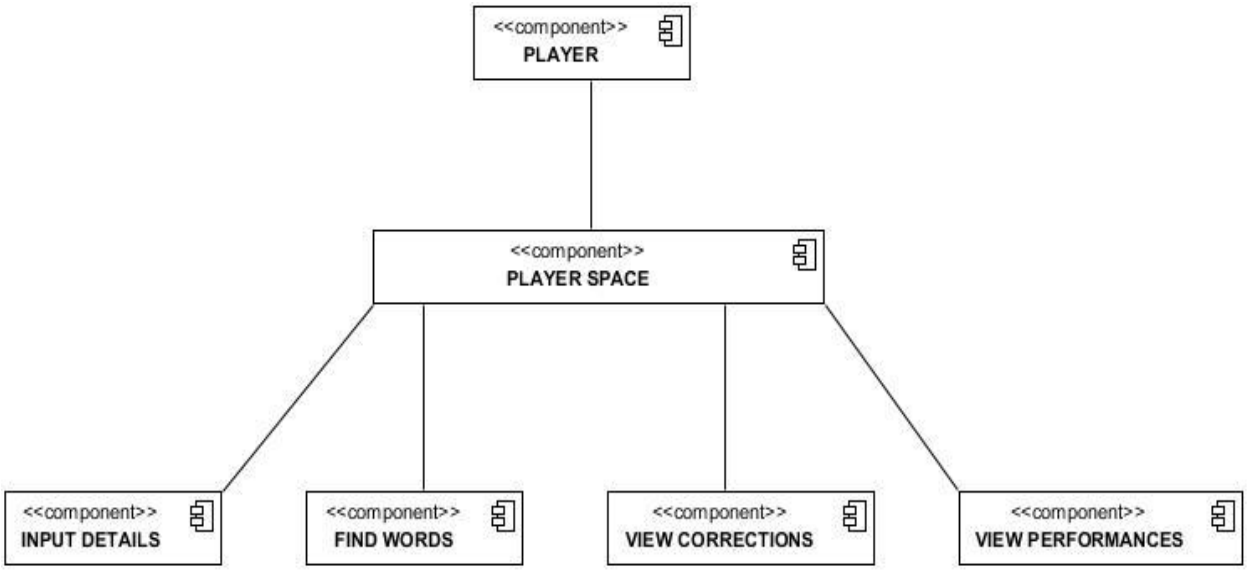


Figure 7: Component Model

V. IMPLEMENTATION

The system was designed using the Top-Down Approach. It makes use of the fundamental program solving techniques. The software is structured in such a way that each subsystem is selected and executed independently. The task is divided into several modules, which come together to give the solution to the problem.

A. SYSTEM RESULT

The home page of the Yoruba-English word search puzzle was designed to contain a form whereby players can select their gender (Male or Female) and age range (i.e. between 0-9, 10-19, 20 and above) using a drop-down menu.

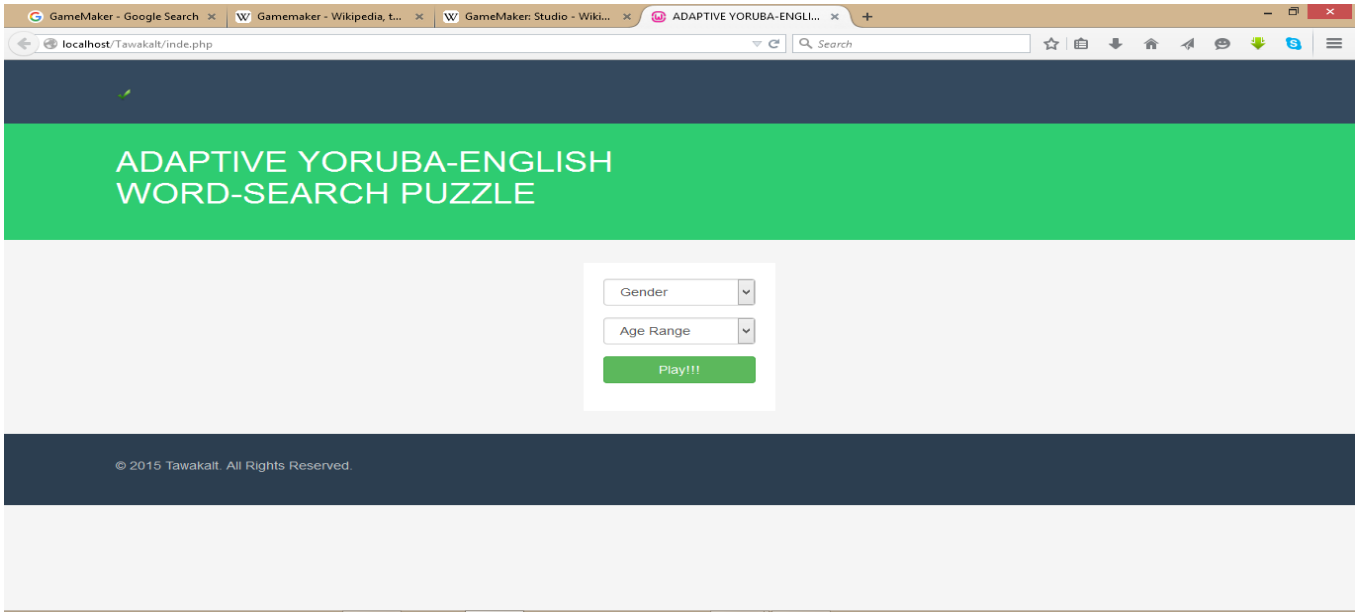


Figure 8: Puzzle Home Page

The puzzle page is solely dependent on the details of the form on the home page. The game is adaptive to the player's gender and age range.

The gender determines the interface of the puzzle while the age range determines the level of Yoruba words to be displayed (2 and 3 letter words for age range 0-9 etc…), and the size of the grid (12*12, 15*15 etc…).

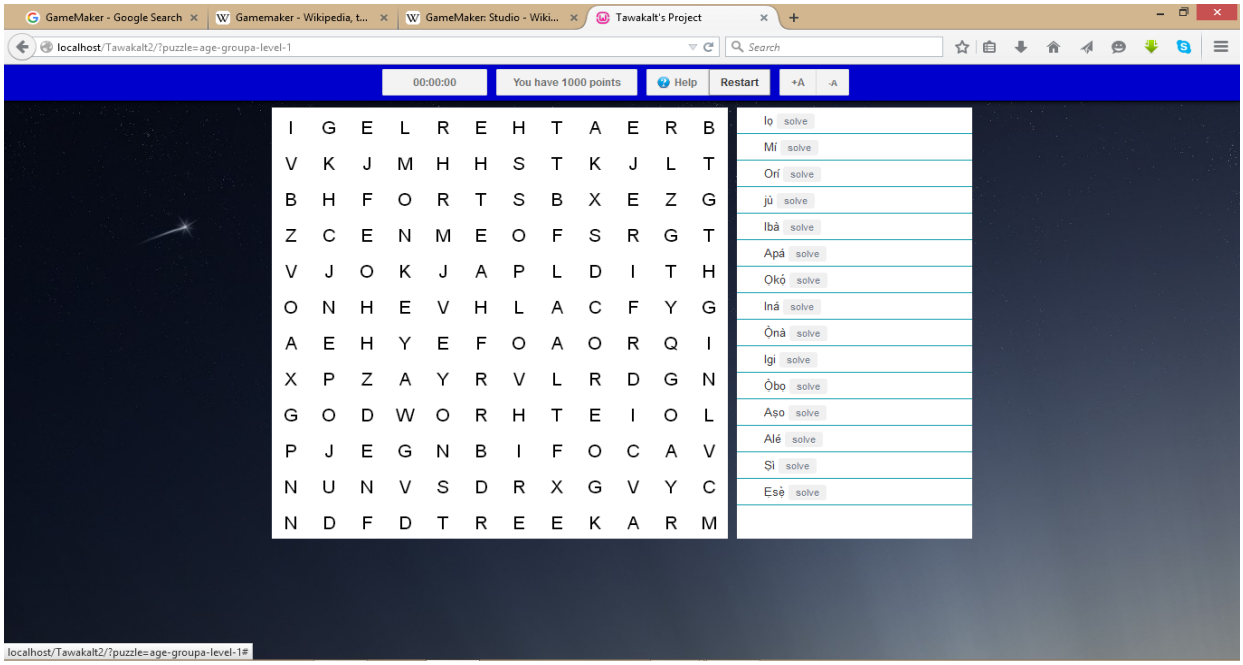


Figure 9: Interface for Male Gender

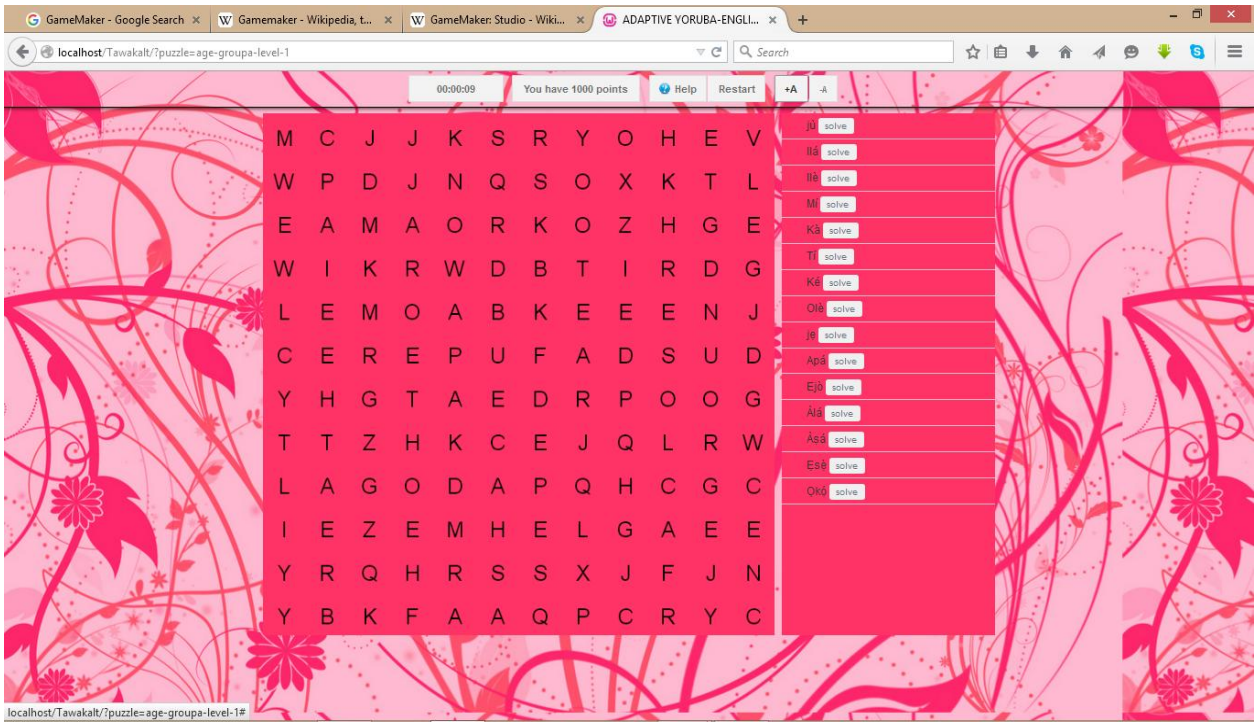


Figure 10: Interface for Female Gender

With the appearance of the puzzle grid, the timer starts counting up with the scores counting down from 1000. All found words appear in red and the corresponding Yoruba words are marked.

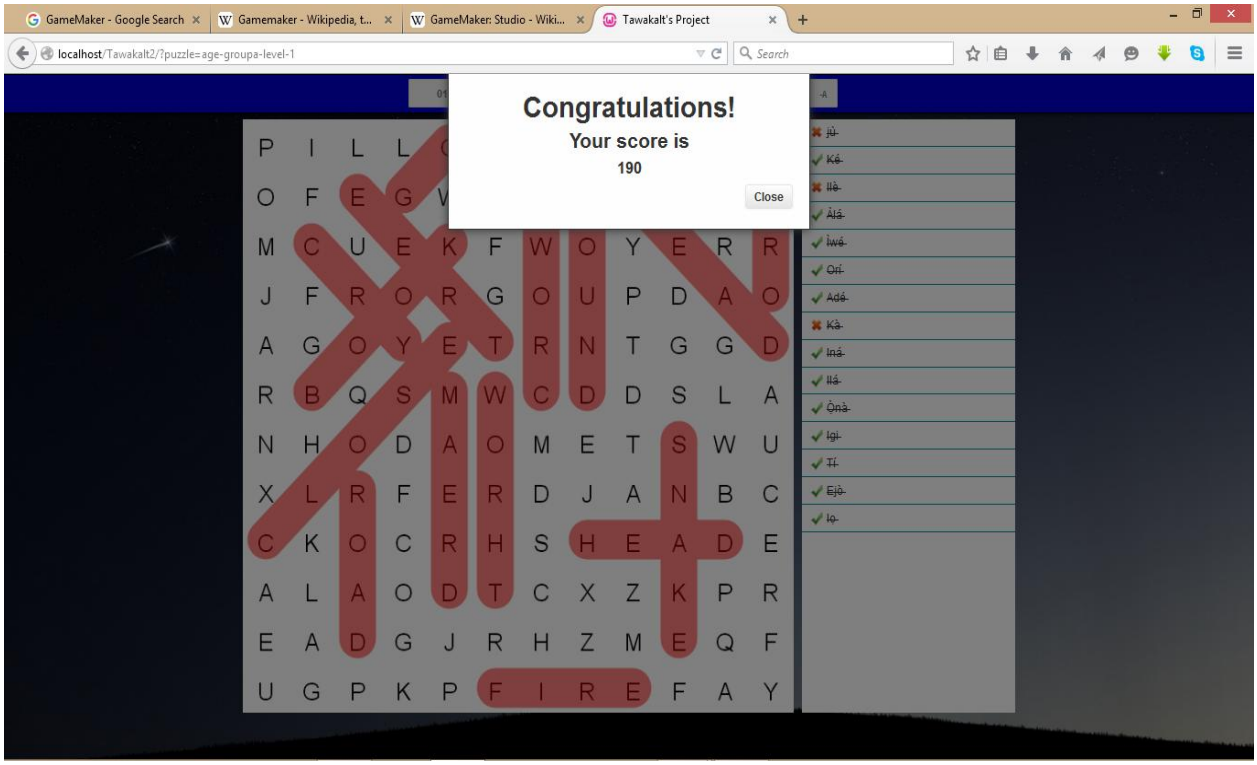


Figure 11: End of Puzzle

At the end of each puzzle, a congratulatory message along with the player's score is shown as a pop-up message.

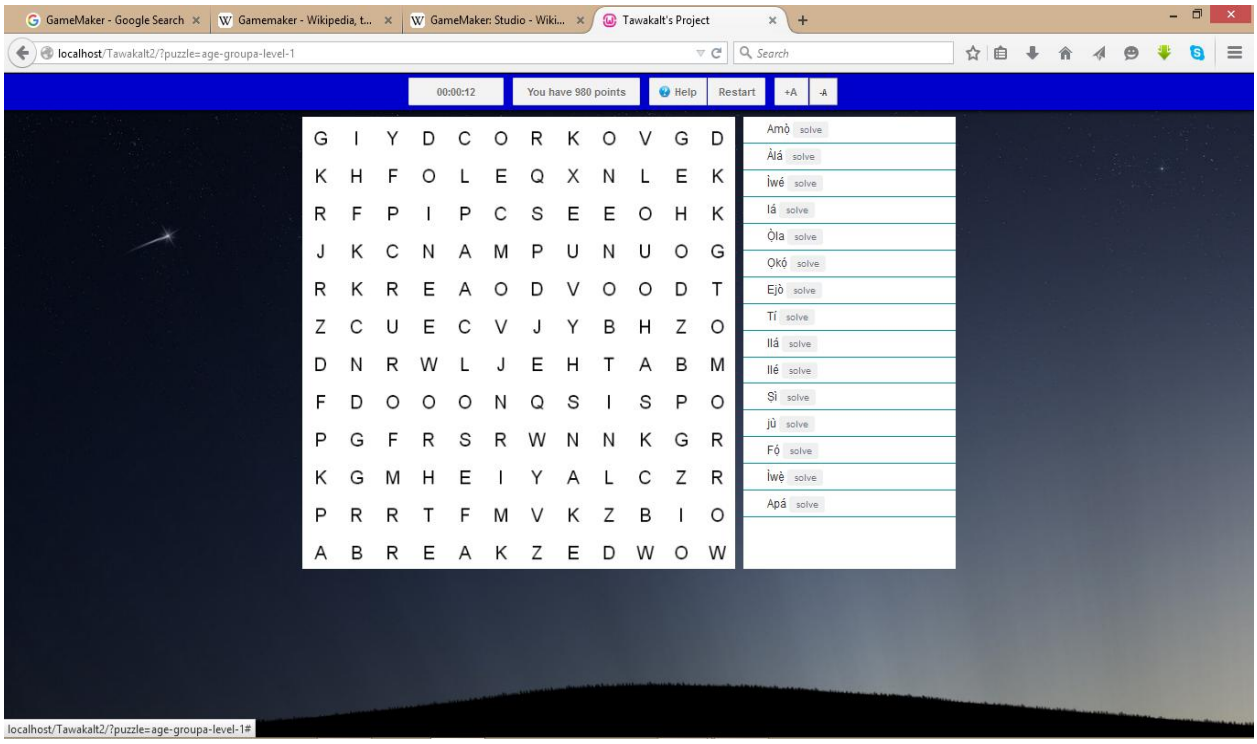


Figure 12: Puzzle Grid 12*12

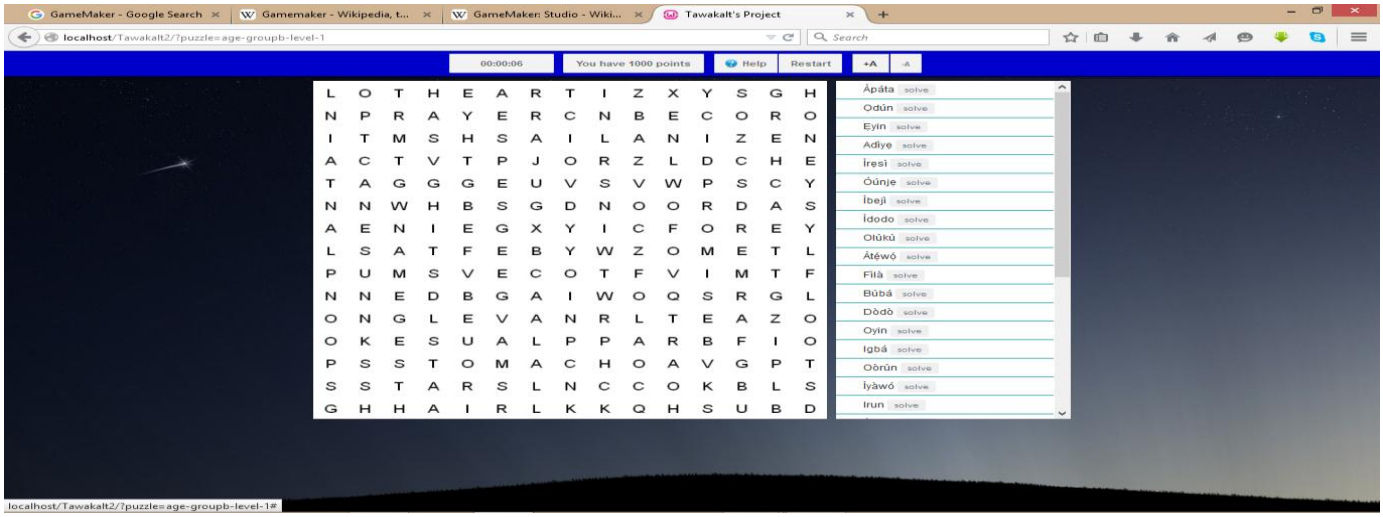


Figure 13: Puzzle Grid 15*15

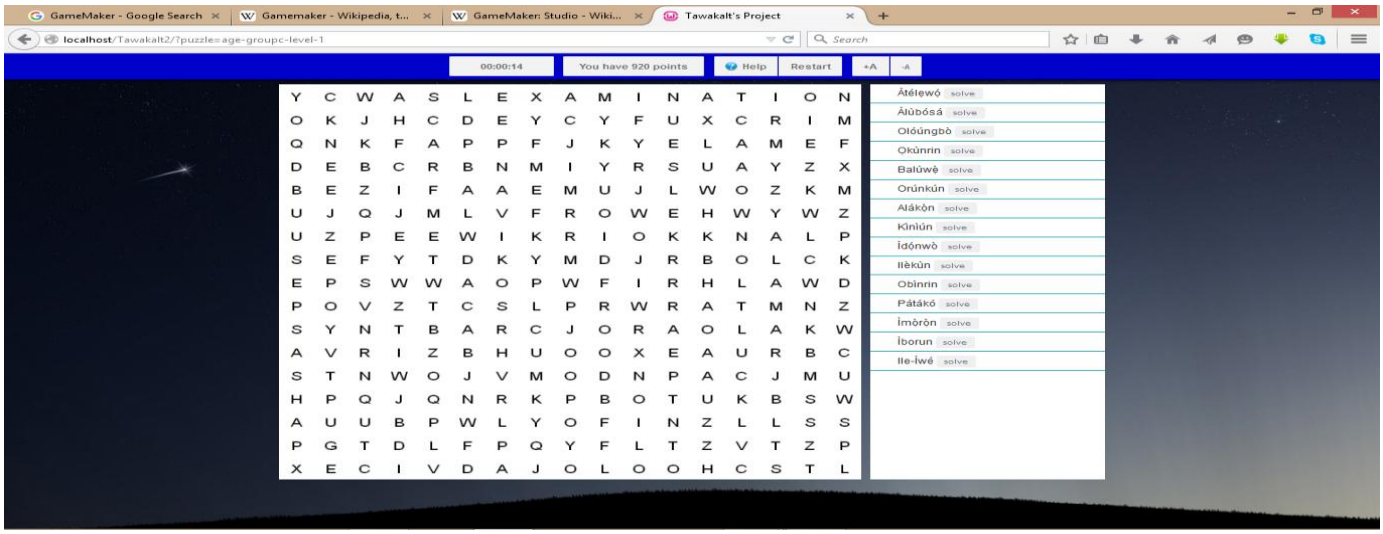


Figure 14: Puzzle Grid 17*17

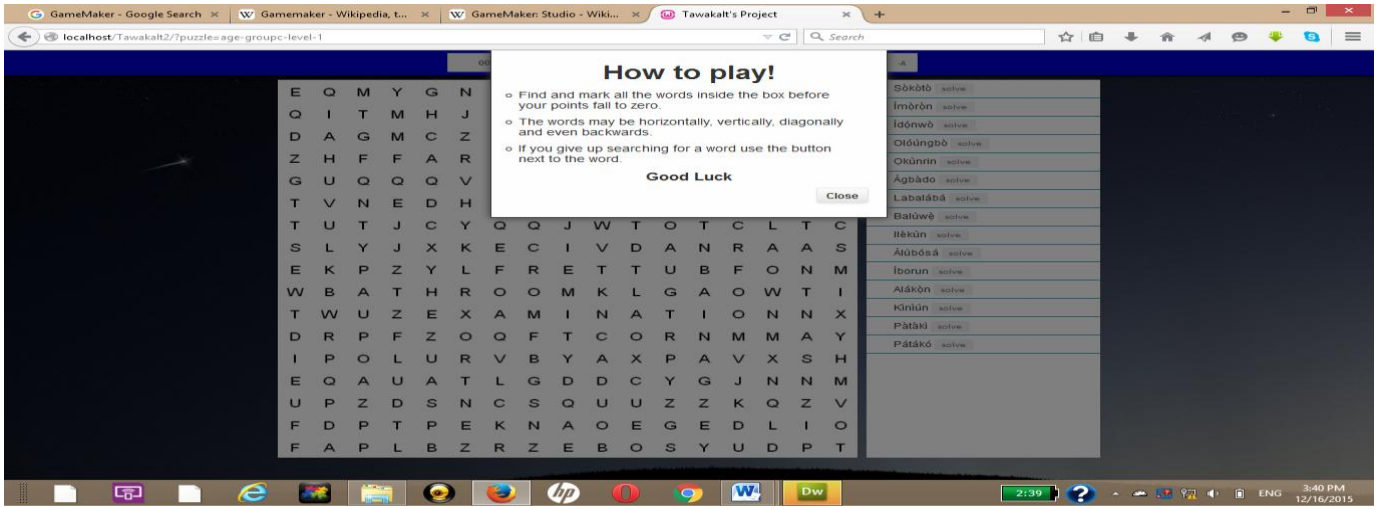


Figure 15: How to Play: Giving instructions on how to play the puzzle

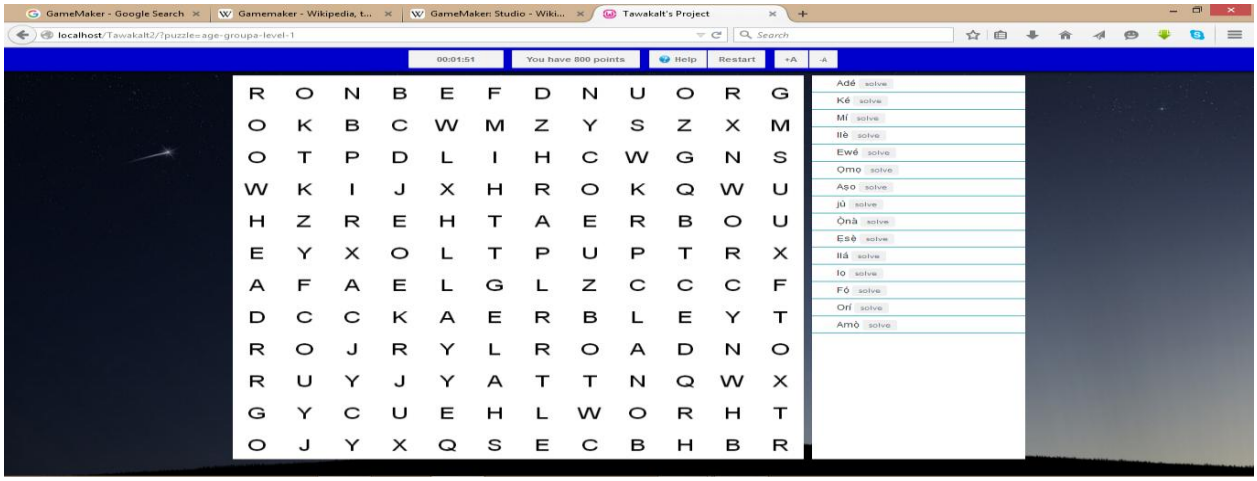


Figure 16: Enlarged Puzzle Grid

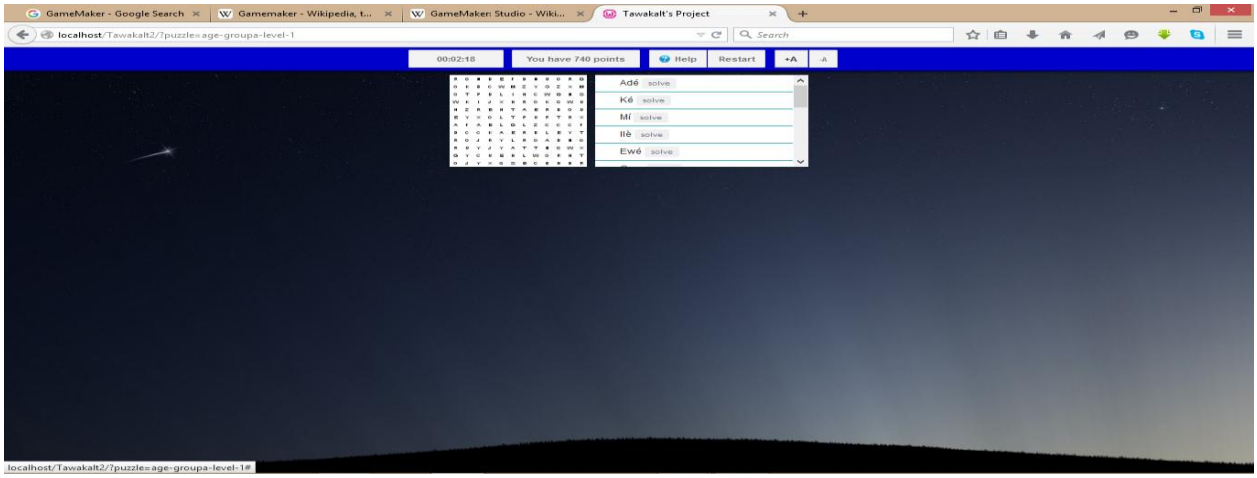


Figure 17: Reduced Puzzle Grid

Words that are difficult to find can be solved with the solve button. Clicking the solve button highlights the English equivalence of the Yoruba word but reduces the player's mark by 50.

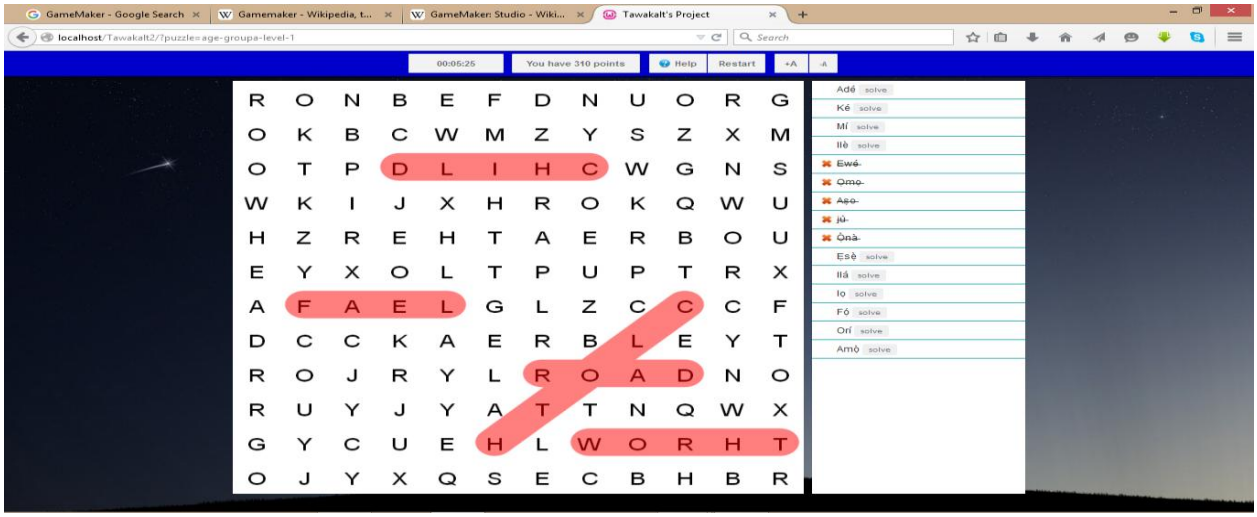


Figure 18: Actions of the Solve Button

VI. CONCLUSION

This paper aimed at applying puzzles and its principles together with the different types of game elements to enhance learning management systems.

Learning is an automatic life-long endeavor for only a selected few. A greater part of learning communities need motivation. Puzzles have proven to be a successful motivation strategy if done the right way. Puzzles may apparently seem like an extrinsic motivation strategy (rewards, badges, points, etc.) but it is much more than that. If puzzle features are placed optimally, they can serve as intrinsic motivators and improve user performance.

VII. RECOMMENDATION

As with any system, software systems always evolve in response to demand for change after deployment. It is impossible to produce a system of any size which do not need to be changed. Once the software is put into use, new requirements emerge and existing requirements change as the business logic and rules running that software changes.

Parts of this software may have to be modified to correct errors that are found in operation, improve its performance or extend its capabilities. This could however come in form of upgrades or new releases and versions. In subsequent upgrades this gamified self-learning system can add some of the following features and processes.

- It can be made available in a mobile version too.
- The scope of the application can be extended by categorizing the words and having players choose their preferred word category.
- Sound can be incorporated to pronounce the Yoruba words.
- Also a new version of this application can add a feedback mechanism where users can raise their complaints and criticisms for the improvement of the system.

VIII. REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[16] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[17] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[18] K. Elissa, "Title of paper if known," unpublished.

[19] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[20] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[21] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

Prediction of Heart Disease Using Techniques in Data Mining: Overview

Aro, T. O
Dept of Computer Science, University of Ilorin, Ilorin, Nigeria
taiwo_aro@yahoo.com

Bajeh, A. O
Dept of Computer Science, University of Ilorin, Ilorin, Nigeria
bajeh_amos@yahoo.com

Olatinwo I. S
Dept of Computer Science, Federal Polytechnic Offa, Kwara State, Nigeria
Suhurat84@yahoo.com

Omokanye, S. O
Dept of Computer Science, University of Ilorin, Ilorin, Nigeria
oladejiomokanye@yahoo.com

ABSTRACT: **With the advancement of data mining applications in healthcare systems, knowledge discovery in database (KDD) of medical data is greatly concerned with extraction of concealed data from huge amount of datasets to provide useful information. Data mining is an important technique that can be used for disease management, diagnosis and prediction in healthcare organizations. This paper discusses different methods and approaches in data mining that have been used to predict heart disease.**

Keywords; Data mining, Heart disease, Diagnosis, Healthcare, Prediction, Knowledge discovery in database.

I. INTRODUCTION

The disease of the heart is one of the major intimidating diseases in the society that has increased mortality rate. Heart disease includes heart attack, heart failure, coronary artery disease and angina [1]. The prediction of heart disease in healthcare has given answers to some difficult questions which help medical experts to arrive at intelligent medical decisions that are not common in the old decision support systems.

Data mining techniques have contributed immensely in transforming large data into specific and more relevant information for knowledge discovery and prediction purpose [2]. Data mining approach in healthcare is a significant component of knowledge discovery in database that is used for the extraction of data associated with several diseases from old dataset in order to facilitate easier prognosis of diseases [3].

Doctors and people suffering from heart ailment need consistent and accurate knowledge of a person's possibility of having a heart disease. Developing an algorithm to estimate the risk of categorizing people with the chance of having heart illness is obtainable through techniques in data mining. The data mining technique can also be engaged to convey knowledge out of dataset in more appropriate form that can be easily understood by people [4].

Techniques in data mining are used to analyze and deduce unknown relationship among features of clinical data to handle some problems such as prediction, diagnosis, control and treatment of diseases [5]. The data mining method is a developing area of great prominence for providing diagnosis and in depth information about medical data [6]. Unlike the statistical methods, data mining approach searches for intriguing information without considering previous postulates on the type of patterns that can be found.

Data mining technology is also referred to as knowledge discovery in database (KDD), it is a domain in which unique and useful information is obtained from huge amount of

dataset [7][8]. The KDD procedure is the mathematical approach of determining pattern types in big database. The most commonly used algorithms in data mining are rule induction, Naïve Bayes, Decision trees, Artificial neural network, K-Nearest Neighbour and Genetic algorithm [1][6].

II. HEART DISEASE

Heart is a fragile part of the human body that must be guarded consciously according to medical experts. Figure 3.1 shows the detail of human heart. Heart disease can be referred to as any form of discomforts that affect the proper working of the human heart. Heart based disease can be classified into different forms. This include:-

(a) Heart Attack: - In heart attack, blood movement to heart is drastically reduced or stopped. Heart attack happens when one or more coronary arteries that take blood to the heart is blocked. Heart attack may lead to a damage of heart muscle.

(b) Heart Failure: This is a situation where the heart refuses to pump enough required blood to the body. This condition does not imply that the heart has stopped functioning. Major signs of heart failure include weight gain, swelling, short breath, cough and fatigue.

(c) Angina: It is a discomfort that happens when the heart fails to get adequate blood. Angina is a symptom of coronary artery disease with pressure, pain in the chest, back, jaws, neck and shoulders.

(d) Arrhythmias: The unusual fastness and slowness that occurs during the heartbeat. Arrhythmias can also be referred to as irregular beat that occurs in the heart.

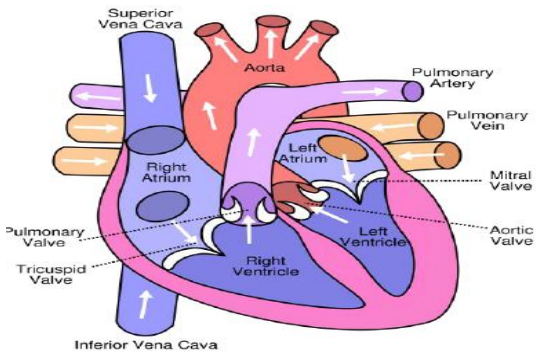


Figure 2.1 Diagram of Human Heart [18]

III.TECHNIQUES IN DATA MINING

Techniques in data mining are methods employed for sorting data in order to identify patterns, these techniques include association, classification, prediction and clustering [4].

(a) Association : looks for patterns based on connection of a particular event to other events. It is commonly used approach for prediction of heart disease as it gives the relationship of unlike features for analysis and grouping out patient with all risk factor needed for prediction [7].

(b) Clustering: - This is unsupervised machine learning in which no class labels are given. It locates and visually documents collections of facts that not previously known.

(c) Classification: - The technique is supervised learning, where the class labels of some training samples are supplied. Mathematical methods such as linear programming, Naïve Bayes, Decision trees and artificial neural network are employed for classification method [6].

(d) Prediction(Forecasting):- determines patterns in data that can result into a reasonable prediction about the future. The prediction technique fits in the prognostic model of data mining [8].

(e) Sequence: searches for patterns where one occurrence results to another future occurrence.

IV ALGORITHMS IN MACHINE LEARNING FOR DATA MINING

These algorithms provide computer systems with ability to learn without completely being programmed [1]. Machine learning is closely related to data mining , the two methods search through data to find patterns. However, machine learning employs data to find patterns in data and then modifies program appropriately instead of data extraction as in human comprehension. Machine learning algorithms in data mining can be classified into:

(a) Supervised learning:- In this method the training data which is the input data has known label. An algorithm is designed to apply what has been learned in the past to new data.

(b) Unsupervised learning: The input data has no labelled. An algorithm is developed by deducing inferences from datasets, this may be either to obtain common rules or by mathematical method to methodically minimize redundancy.

(c)Semi-supervised: The input data comprises of labelled and unlabeled data. Here, there is a need to predict but the algorithm must learn the structures to arrange the data properly in order to have accurate prediction.

V. RELATED WORK

Data mining techniques have been employed to conduct the prediction and diagnosis of heart based disease in the

medical domain. This section reviews several researchers work using these techniques.

[8] developed an enhanced system for prediction of heart disease using classification techniques in data mining. They evaluated the prediction systems by employing more number of input features. The model used medical parameters such as blood, cholesterol, sex, pressure to forecast the chance of a person having heart ailment. Two features which include obesity and smoking were added. The heart disease database were analyzed using Naïve Bayes, Decision trees and Artificial neural network. The performance evalaution of these techniques compared according to their accuracy. The results gave precision of 100% for neural network, 99.62% for decision trees and 90.74% for naïve bayes.

[6] conducted general review on prediction of cardiovascular heart hazard analysis and evaluation performance using several techniques of data mining. The study reviewed various performance of algorithms, approaches and results for prediction of heart disease applying data mining. Evaluation, result methods and a summary of findings were discussed. A conclusion was reached in his study that data mining techniques can offer a reliable performance for heart disease prediction.

[5] used a fuzzy system and K-nearest neighbor classifier to perform prediction of heart disease. The study integrated voting with KNN to enhance accuracy in the diagnosis task. The KNN algorithm proves to be more efficient when implemented with fuzzy rules for certain set of disease. He also came up with an observation that KNN works better when quality datasets is used.

[9] genetic algorithm and KNN method were considered to improve heart disease classification accuracy. Genetic algorithm was employed to reduce irrelevant and redundant data attributes and also select the attributes that contribute more towards classification. KNN classifier was trained to categorize heart disease based on dataset as either healthy or sick. The performance of the data mining approach for heart disease prediction was carried out with 6 medical datasets and 1 non-medical dataset. The results obtained reviewed that integrating KNN and genetic algorithm will improve the classification accuracy for many datasets.

[10] developed a system for coronary heart disease prediction using neuro-fuzzy integrated approach two level. They introduced two layered approaches for detecting the disease possibility. The critical issues that is compulsory for occurrence of coronary heart disease were considered at first level and the next one obtained at second level. The two phase method has significant effect on the performance of their system as it contributes to the prediction of disease chances accurately. UCI machine learning heart disease database was used to train the neural network and then use fuzzy rules to forecast levels of heart disease as low, moderate or critical.

[11] designed heart disease diagnosis with classification model. They divided the publicly available heart disease database using K-means method where K = 2 values on classes cluster evaluation of testing rule. Classification rule methods of data mining, which is also called projective adaptive resonance theory analyzed on clustered relevant dataset. 10 fold cross validation method was considered to measure unbiased total estimation of the prediction system. The precision of K-means clustering, PART and PART-based K-means clustering are 81.08%, 79.05% and 84.12% correspondingly. The result of three classification methods reviewed that classification-base clustering predicts cardiovascular disease of heart with better precision.

[12] came up with a web-based disease of heart decision support system using classification modelling methods of data mining. They used Naïve Bayes, Decision trees and Neural Network for classification. The decision support system used six phases of CRIP-DM methodology to develop data mining sysgtems. Data mining extension, SQL-Style query language for data mining, graphical visualizations and tabular were combined to improve analysis of their experimental results. The decision heart disease support system results indicated that each method of classification has its distinct power in realizing the plans of the defined mining aims.

[13] proposed a model for the prediction of risk level of heart ailment. They compared the performance two modelling methods; Naïve Bayes and Genetic Algorithm. Genetic algorithm was mentioned to reduce set of features or factors that contribute more to heart ailment. A conditional probability method known as Naïve Bayes method was applied on antique heart disease database to produce relationships that exist among the factors. The comparative analysis showed that two modelling data mining methods not accurate for prediction of intensity risk level of heart disease. They planned to develop an intelligent decision support system by combining Apriori, Genetic and Fuzzy methods in order to build optimal prediction system for risk level of heart disease.

[14] developed prediction model of heart disease using genetic algorithm method and associative classification. They located Gini index of attribute. Attribute with minimum Gini index selected for class association rule generation. Evaluation of fitness of rule was done with z statistics. After the rule evaluation, rules with very high fitness were stored in a database. They applied genetic functions on these rules. Classifiers were constructed based on the rules generated.

Results showed that most of the classifier rules aid the optimal prediction of heart ailment.

[15] presented a study on heart disease diagnosis using predictive data mining method. The prediction and diagnosis system for heart disease was based predictive mining. Naives Bayes and Decision tree were used and compared with each other. UCI database of 294 records with 13 attributes was used as source data. The experimental result showed that Naives Bayes perform better than Decision tree.

[16] used cluster-based association rule mining for prediction of heart attack. The model combined the concept of sequence numbers and clustering. The system transforms medical data into binary. The cluster based association rule was applied on transactional data. Experimental results showed that the model outperformed other algorithms.

[17] presented a study on the application of K-Nearest Neighbour (KNN) for heart disease diagnosis. They used KNN on Cleveland Heart Disease database to investigate its efficiency in the prediction of heart disease. Voting was combined with KNN to improve precision. The experimental results showed that application of KNN achieved higher precision than Neural Network in heart disease diagnosis. The system also reviewed that applying voting did not improve the KNN precision for the diagnoses of heart disease.

[18] developed a predictive system for heart disease prediction using methods of data mining. They employed Neural Network, frequent item set generation and K-means clustering as data mining techniques with Apriori techniques to predict whether a person suffers from heart disease or not. Medical profiles such as age, blood sugar, blood pressure and sex were also considered as parameters to predict the chance of a person getting heart disease. These techniques performance were compared through sensitivity, specificity and accuracy. Result reviewed that Artificial Neural Networks performed better than K-means clustering in all the parameters.

Several studies have been proposed for treatment and prediction of heart disease, but most algorithms designed so far only engaged directly common data mining techniques which are Decision tree, Artificial Neural Network, K-Nearest Neighbour and Support Vector Machines. Little attention has been given towards irrelevant or redundant data that exist in the datasets. Most algorithms used in the past employed convectional methods for the reduction of data like Principal Component Analysis or Linear Discriminant Analysis, these methods are still very high in computation. In order to develop a robust heart disease diagnosis, it requires to introduce more of the effective data reduction techniques such as nature inspired optimization algorithms like particle swarm optimization, genetic algorithm and Tabu search algorith to select most relevant data from heart disease datasets. This will contribute towards the reduction in complexity that may occur during prediction which in turn allow a very fast and accurate heart predictive system.

VI. CONCLUSION

Heart disease is life threaten disease which occurs due to obstruction or narrowing of coronary arteries. This disease has claimed so many lives in the society. The prediction and diagnosis of heart disease is very important in order to assist individual to commence earlier treatment before it gets to critical stage. This paper discusses some methods used in data mining for prediction of heart disease. Also futher areas to concentrate on in order to develop a robust predictive are mentioned.

REFERENCES

[1] D. Chandna, "Diagnosis of Heart Disease Using Data Mining Algorithm," *Inetrnational Comput. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 1678–1680, 2014.

[2] O. O. Adeyemo and T. O. Adeyeye, "Comparative Study of ID3 / C4 . 5 Decision tree and Multilayer Perceptron Algorithms for the Prediction of Typhoid Fever," *African J. Comput. ICT*, vol. 8, no. 1, pp. 103–112, 2015.

[3] A. S, Aslam and I, "Data Mining Algorithms and their applications in Education Data Mining," *Int. J. Adv. Res. Comput. Sci. Mnangement Stud.*, vol. 2, no. 7, pp. 50–56, 2014.

[4] M. A. Khaleel, S. K. Pradhan, and G. N. Dash, "Finding Locally Frequent Diseases Using Modified Apriori Algorithm," *Int. J. Adv. Res. Comput. Coomunication Eng.*, vol. 2, no. 10, pp. 3792–3797, 2013.

[5] A. A. Chaudhari, "Fuzzy & Datamining based Disease Prediction Using K-NN Algorithm," *Int. J. Innov. Engineeering Technol.*, vol. 3, no. 4, pp. 9–14, 2014.

[6] D. C. Bindushree, "Prediction of Cardiovascular Risk Analysis and Performance Evaluation using various Data mining Techniques : A Review," *Int. J. Enginnering Res.*, vol. 5013, no. 5, pp. 796–800, 2016.

[7] T. Karthikeyan, B. Ragavan, and V. A. Kanimozhi, "A Study on Data mining Classification Algorithms in Heart Disease Prediction," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 5, no. 4, pp. 1076–1081, 2016.

[8] S. S. Dangare, C. S and Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques," *Int. J. Comput. Appl.*, vol. 47, no. 10, pp. 44–48, 2012.

[9] M. Akhil, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," *Procedia Technol.*, vol. 10, pp. 85–94, 2013.

[10] A. K. Sen, S. B. Patel, and D. P. Shukla, "A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level," *Int. J. Eng. Comput. Sci.*, vol. 2, no. 9, pp. 1663–1671, 2013.

[11] A. Kumar, P. Prabhat, and K. L. Jaiswal, "Classification Model for the Heart Disease," *Glob. J. Med. Resarch Dis.*, vol. 14, no. 1, 2014.

[12] R. Palaniappan, S and Awang, "Web-Based Heart Disease Decision Support System using Data Mining Classification Modeling Techniques," *Proc. ii AS2007*, pp. 157–167, 2007.

[13] S. Ratnakar, K. Rajeswari, and R. Jacob, "Prediction of Heart Disease using Genetic Algorithm," *Int. J. Adv. Comput. Eng. Neworking*, vol. 1, no. 2, pp. 51–55, 2013.

[14] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Heart Disease Prediction System using Associative Classification and Genetic Algorithm," *Int. Conf. Emerg. Trends Electr. Electron. Commun.*, pp. 183–192, 2013.

[15] B. Venkatalakshmi and M. V Shivsankar, "Heart Disease Diagnosis Using Predictive Data mining," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 3, no. 3, pp. 1873–1877, 2014.

[16] M. Jabbar, P. Chandra Dr., and B. L. Deekshatulu, "Cluster based Association rule mining for Heart Attack Prediction," *J. Theor. Appl. Inf. Technol.*, vol. 32, no. 2, pp. 196–201, 2011.

[17] M. Shouman, T. Turner, and R. Stocker, " Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients," *Int. J. Inf. Educ. Technol.*, vol. 2, no. 3, pp. 220–223, 2012.

[18] A. D. Souza, "Heart Disease Prediction Using Data Mining Techniques," *Int. J. Res. Eng. Sci.*, vol. 3, no. 3, pp. 74–77, 2015.

Boundedness of Genetic Codes and its Applications

Ogundeji Olumide Ayobami
Department of Computer Science,
University of Ilorin,
Nigeria.
ogundeji.olu@gmail.com

Oluwade Bamidele A.
Department of Computer Science,
University of Ilorin
Nigeria.
deleoluwade@yahoo.com

Abstract: **The recent development of bioinformatics begins with research on genes and moves to the molecular sequence, then to molecular conformation, from structure to function, from systems biology to network biology, and further investigates the interactions and relationships among, genes, proteins, and structures. Mathematics can be used to understand life from the molecular to the biosphere levels, including the origin and evolution of organisms, the nature of genomic blueprints, and the universal genetic code as well as ecological relationships. Mathematics and biological data have a synergistic relationship. Biological information creates interesting problems, mathematical theory and methods provide models for understanding them and biology validates the mathematical models. This work uses mathematical theorems to check the boundedness of genetic codes and to form a new theorem that is based on genetic codes.**

Keywords: sequence, boundedness, genetic code, amino acid, proteins, genes

I. INTRODUCTION

Bioinformatics has evolved in recent times with research on genes and moves to the molecular sequence, then to molecular conformation, from structure to function, from systems biology to network biology, and further checks the interactions and relationships among, genes, proteins, and structures.

This new reverse model sets a theoretical starting point for a biological investigation. It sets a new line of investigation with a unifying principle and uses mathematical tools extensively to clarify the ever - changing phenomena of life quantitatively and analytically. It is well known that there is more to life than the genomic blueprint of each organism. Life functions within the natural laws that we know and those that we do not know. Life is founded on mathematical patterns of the physical world. Genetics exploits and organizes these patterns.

Life is founded on mathematical patterns of the physical world. Genetics exploits and organizes these patterns. Mathematics helps us understand how monomers become polymers necessary for the assembly of cells. Mathematics can be used to understand life from the molecular to the biosphere levels, including the origin and evolution of organisms, the nature of genomic blueprints, and the universal genetic code as well as ecological relationships. Mathematics and biological data have a synergistic relationship. Biological information creates interesting problems, mathematical theory and methods provide models for understanding them, and biology validates the mathematical models. A model is a representation of a real system. Real systems are too complicated, and observation may change the real system. A good system model should be simple, yet powerful enough to capture the behaviour of the real system. Models are especially useful in bioinformatics. [7].

[14] examined a scenario in which evolution drives the emergence of a genetic code by selecting for an amino acid map that minimizes the impact of errors. Also, treated the stochastic mapping of codons to amino acids as a noisy information channel with a natural fitness measure. Organisms compete by the fitness of their codes and, as a result, a genetic code emerges at a supercritical transition in the noisy channel, when the mapping of codons to amino acids becomes non-random. At the phase transition, a small expansion is valid and the emergent code is governed by smooth modes of the Laplacian of errors.

II. DEFINITION OF SOME TECHNICAL TERMS

Boundedness: A sequence is called *bounded from above* if there exists , which satisfies the inequality for all . In this case, we say M is an *upper bound* for x. we say is *bounded from below* if there exists which satisfies the inequality for all . In this case, we say that m is a *lower bound* of . Therefore, any sequence that is bounded from above and below is bounded. [12]

Genetic Code: The genetic code is the full set of relationships between codons and amino acids (or stop signals) is known as the genetic code. [5]

The genetic code is a set of rules defining how the four-letter code of DNA is translated into the 20-letter code of amino acids, which are the building blocks of proteins. The genetic code is a set of three-letter combinations of nucleotides called codons, each of which corresponds to a specific amino acid or stop signal.[9].

Sequence Alignment: The result of a comparison of two or more gene or protein sequences in order to determine their degree of base or amino acid similarity. Sequence alignments are used to determine the similarity, homology, function or other degree of relatedness between two or more genes or gene products. [8]

III. THE GENETIC CODE

The canonical genetic code maps 64 codons onto 21 messages –20 amino-acids and a stop-signal. (Every codon encodes a single message and each message is encoded by at least one codon: a 64-21onto mapping). All extant living organisms use this code, or minor variations thereof [6,7], to synthesize the proteins encoded by their genomes and this strongly suggests that all modern life evolved from a last universal common ancestor (LUCA) of more than 3.5 billion years ago. The code shows pattern that are similar [2,17]. The code would have evolved in a pre-LUCA RNA-world of life that, at least initially, was incapable of directed protein synthesis [1,3]. Most messages are encoded by several synonymous codons (the degeneracy of the code). [13].

The genetic code was finally "cracked" in1966. Marshall *et al* demonstrated that a sequence of three nucleotide bases, a codon or triplet, determines each of the 20 amino acids found in nature. This means that there are 64 possible combinations (43 = 64) for 20 amino acids. They formed synthetic messenger ribonucleic acid (mRNA) by mixing the nucleotides of RNA with a special enzyme called polynucleotide phosphorylase. This resulted in the formation of a single - stranded RNA in this reaction. The question was how these 64genetic codes could code for 20 different amino acids. [11] synthesized poly(U) by reacting only uracil nucleotides with the RNA - synthesizing enzyme, producing – UUUU – . They mixed this poly(U) with the protein - synthesizing machinery of *Escherichia coli* in vitro and observed the formation of a protein. This protein turned out to be a polypeptide of phenylalanine. They showed that a triplet of uracil must code for phenylalanine. Philip Leder and Nirenberg found an even better experimental protocol to solve this fundamental problem. By 1965 the genetic code was solved almost completely. They found that the "extra" codons are merely redundant: Some amino acids have one or two codons, some have four, and some have six. Three codons (called *stop codons*) serve as stop signs for RNA – synthesizing proteins.

A *DNA* or *genetic sequence* is a series of letters signifying the major structure of a real or hypothetical DNA molecule or strand, with the ability to carry information. The possible letters are A, C, G, and T, representing the four nucleotide subunits of a DNA strand: adenine, cytosine, guanine, and thymine bases covalently linked to a phospho - backbone. In the typical case, the sequences are printed abutting one another without gaps, as in the sequence AAAGTCTGAC, going from 5′ to 3′ from left to right. A succession of any number of nucleotides greater than four is liable to be called a sequence. With regard to its biological function, which may depend on context, a sequence may be *sense* or *antisense*, and either coding or noncoding. DNA sequences may also contain "junk DNA." In some special cases, letters other than A, T, C, and G are present in a sequence. These letters represent ambiguity. The rules of the International Union of Pure and Applied Chemistry (IUPAC) are as follows:

A = adenine
C = cytosine
G = guanine
T = Thymine
R = GA (purine)
Y = TC (pyrimidine)
K = GT (keto)
M = AC (amino)
S = GC (strong bonds)
W = AT (weak bonds)
B = GTC (all but A)
D = GAT (all but C)
H = ACT (all but G)
V = GCA (all but T)
N = AGCT (any)

A peptide or *amino acid sequence* is the order in which amino acid residues, connected by peptide bonds, lie in a chain in peptides and proteins. The sequence is generally reported from the N - terminal end, which contains a free amino group, to the C - terminal end, which contains a free carboxyl group. A peptide sequence is often called a *protein sequence* if it represents the primary structure of a protein. Amino acids are the basic structural building units of proteins. They form short polymer chains called *peptides* or longer chains called either *polypeptides* or *proteins*. The process of such formation from an mRNA template, known as *translation*, is part of protein biosynthesis. Twenty amino acids are encoded by the standard genetic code (Table 1). Proteins are defined by their unique sequence of amino acid residues; this sequence is the primary structure of the protein. Just as the letters of the alphabet can be combined to form an almost endless variety of words, amino acids can be linked in varying sequences to form a vast variety of proteins.

According to [16] in his work argued that the first information system emerged on the earth as primordial version of the genetic code and genetic texts. The natural appearance of arithmetic power in such a linguistic milieu is theoretically possible and practical for producing information systems of extremely high efficiency. In this case, the arithmetic symbols should be incorporated into an alphabet, i.e. the genetic code

Table 1 **Amino Acid with their Abbreviations Source: [10]**

Amino acid	Three-Lettered Abbreviation	One-Lettered Abbreviation
Alanine	Ala	A

Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid (Aspartate)	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid (Glutamate)	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Aspartic acid or Asparagine	Asx	B
Glutamine or Glutamic acid.	Glx	Z
Any amino acid.	Xaa	X
Termination codon		TERM

IV. BOUNDEDNESS OF A SEQUENCE

Definition 1: A *sequence* is a function whose domain of definition is the set $\mathbb{N}$ of all natural numbers. Sequences obtain different names with respect to their range. If the range of the sequence is $\mathbb{R}$, then we call this sequence a real number sequence (or real sequence). If the terms are all rational numbers, then, we called this sequence rational number sequence (or rational sequence). [12]. Generally, we use the notation

$$x = \{x_k\}_{k=1}^{\infty} \forall k \in \mathbb{N}$$

To represent sequences. For each value of k, the term x_k is known as k^{th} term of x. The space of all sequences denoted by ω.

The above definition can be re-written informally as follows:

A sequence $x = \{x_k\}_{k=1}^{\infty}$ of real numbers is a function $x : D(x) \subset \mathbb{N} \to \mathbb{R}$ of an infinite subset (x) of the natural numbers $\mathbb{N}$ into the real numbers $\mathbb{R}$ defined by

$$x(k) = x_k \in \mathbb{R} \forall k \in \mathbb{N}$$

Definition 2: Let $\{x_k\}$ be a sequence and let $\{k_n\}$ be a strictly increasing sequence of natural numbers. The sequence $\{x_{k_n}\}$ is called a *subsequence* of $\{x_k\}$.

Definition 3: A sequence $\{x_k\}$ is called *bounded from above* if there exists $M \in \mathbb{R}$, which satisfies the inequality $x_k \leq M$ for all $k \in \mathbb{N}$. In this case, we say M is an *upper bound* for x.

Definition 4: we say $\{x_k\}$ is *bounded from below* if there exists $m \in \mathbb{R}$ which satisfies the inequality $m \leq x_k$ for all $k \in \mathbb{N}$. In this case we say that m is a *lower bound* of x.

Definition 5: We say that a sequence $\{x_k\}$ is bounded if there exists a real constant $U > 0$, which satisfies the inequality

$$|x_k| \leq U \forall k \in \mathbb{N}$$

V. DISCUSSION

Definition 6: A function is an ordered triple $\langle f, A, B \rangle$ such that

1. A and B are sets, and $f \subseteq A \times B$,
2. For every $x \in A$ there is some $y \in B$ such that $\langle x, y \rangle \in f$
3. If $\langle x, y \rangle \in f$ and $\langle x, z \rangle \in f$, then $y = z$; in other words, the assignment is unique in the sense that an $x \in A$ is assigned at most one element of B. A Is called the domain of f, and B its codomain.

It is customary to write the function $\langle f, A, B \rangle$ as $f : A \to B$. Also, if $\langle x, y \rangle \in f$, then we will usually write $y = f(x)$, and call y the image of x under f. [4]

Example 1: A function $f : A \to B$ is invertible if and only if f is both one-to-one and onto.

If $f : A \to B$ is one-to-one and onto, then f is called a one-to-one correspondence between A and B. This terminology comes from the fact that each element of A will then correspond to a unique element of B and vice versa. Some texts use the terms injective for a one-to-one function, surjective for an onto function, and bijective for a one-to-one correspondence. [14]. In table 2, the fist row and first column are only used to generate a two letter code. In Table 3, the first row and first column are only used to generate a three letter code (CODON).

Table 2 Nucleotides to generate the genetic code

	A	G	T	C
A	AA	AG	AT	AC
G	GA	GG	GT	GC
T	TA	TG	TT	TC
C	CA	CG	CT	CC

Table 3 The Genetic Codes

	A	G	T	C
AA	AAA	AAG	AAT	AAC
AG	AGA	AGG	AGT	AGC
AT	ATA	ATG	ATT	ATC
AC	ACA	ACG	ACT	ACC
GA	GAA	GAG	GAT	GAC
GG	GGA	GGG	GGT	GGC
GT	GTA	GTG	GTT	GTC
GC	GCA	GCG	GCT	GCC
TA	TAA	TAG	TAT	TAC
TG	TGA	TGG	TGT	TGC
TT	TTA	TTG	TTT	TTC
TC	TCA	TCG	TCT	TCC
CA	CAA	CAG	CAT	CAC
CG	CGA	CGG	CGT	CGC
CT	CTA	CTG	CTT	CTC
CC	CCA	CCG	CCT	CCC

Using one to one correspondence (1-1) or bijective we have:

$$A \to 1$$
$$G \to 2$$
$$T \to 3$$
$$C \to 4$$

Table 4 Number Equvalence of the Genetic Code

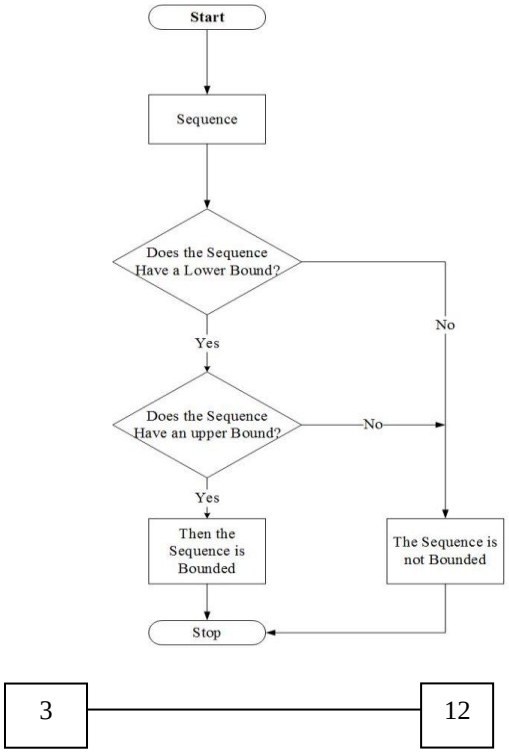
	1	2	3	4
11	111	112	113	114
12	121	122	123	124
13	131	132	133	134
14	141	142	143	144
21	211	212	213	214
22	221	222	223	224
23	231	232	233	234
24	241	242	243	244
31	311	312	313	314
32	321	322	323	324
33	331	332	333	334
34	341	342	343	344
41	411	412	413	414
42	421	422	423	424
43	431	432	433	434
44	441	442	443	444

As usual, the first row and first column merely correspond to the first rwo and first column of tabe 3.

Table 5: Addtion of Each Entry of the Non leading Row (First Row) and Non Leading Colun (First Column) of Table 4

3	4	5	6
4	5	6	7
5	6	7	8
6	7	8	9
4	5	6	7
5	6	7	8
6	7	8	9
7	8	9	10
5	6	7	8
6	7	8	9
7	8	9	10
8	9	10	11
6	7	8	9
7	8	9	10
8	9	10	11
9	10	11	12

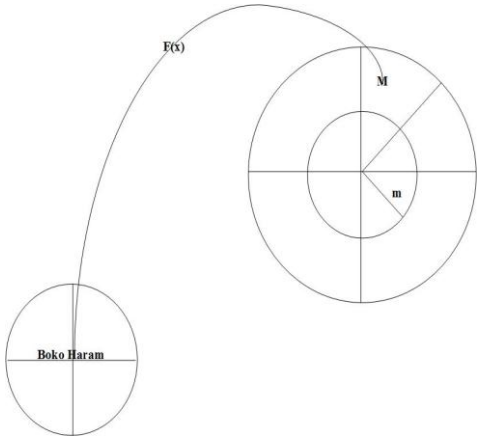
VI. FLOW CHAT



From **Definition 4**, we say 3 is the lower bound for the genetic code which means that and . This implies that 3 is the Greatest Lower Bound (GLB) for the sequence.

Also from **Definition 3**, we say 12 is the upper bound for the genetic code which means that and . This implies that 12 is the Least Upper Bound (LUB) for the sequence.

VII. APPLICATIONS



$$m \leq f(x) \leq M$$

Before any bombing there must be a bound radius of the spot where the bombing will take place.

Also in mining of either gold or crude oil there must be a specific location which will have a lower and upper bound, then it will make the mining more easy and exact.

Also, money to buy goods must be quantified, if it is not then it is not bounded.

VIII. CONCLUSION

In conclusion, we can see that the genetic code can be transformed to mathematical form by using the concept of one-to-one correspondence (bijective) and we analysed these genetic codes using mathematical theorems to show that the genetic code is bounded.

REFERENCES

[1] Atkins, J.F., Gesteland, R.F., Cech, R. (Eds.), RNA Worlds. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York*,. 2011

[2] Crick, F.H.C., The origin of the genetic code. *J. Mol. Biol.*, 38, 367–379. 1968

[3] Deamer, D., Szostak, J.W. (Eds.), The Origins of Life. *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York*.' 2010

[4] Ivo Düntsch and Günther Gediga, Sets, Relations, Functions,First Edition, Methoδos Publishers (UK) pp 35-39, 2000

[5] Khan Academy, The genetic code. In *Biomolecules*. Retrieved from https://www.khanacademy.org/test-prep/mcat/biomolecules/dna/v/the-genetic-code. 2016

[6] Knight, R.D., Freeland, S.J., Landweber, L.F., Rewiring the keyboard: evolv-abilitity of the genetic code. *Nat. Rev. Genet. 2*, 49– 58. 2001

[7] Koonin, E.V., Novozhilov, A.S., Origin and evolution of the genetic code: *the universal enigma. Life 61*, 99– 111. 2009

[8] Matthew Heand Sergey Petoukhov, Mathematics of Bioinformatics: Theory, Methods and Applications, John Wiley & Sons, Inc. 2011

[9] Nature Genetic Code: http://www.nature.com/scitable/definition/genetic-code-13, (accessed on 14/06/2016). 2016

[10] NBCI: Amino Acids Abbreviations IUPAC. http://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioRevie w/iupac_aa abbrev-iations.html (accessed on 13/06/2016). 2016

[11] Nirenberg J. H. and Matthaei J. H., The Genetic Code 11. Scientific American,(March) vol. 190:80-94, 1963

[12] Olubummo, A., Introduction to Real Analysis. *HEBN Publisher, Ibadan*., 1979

[13] Reijer, L., The Graph, Geometry and Symmetries of the Genetic Code with Hamming Metric. *Symmetry*, www.mdpi.com/journal/symmetry. 2015

[14] Seymour Lipschutz and Marc Lars Lipson, Theory and Problems of Discrete Mathematics, Third Edition, The McGraw-Hill Companies, 43-46, 2007

[15] Tsvi Tlusty, , A model for the emergence of the genetic code as a transitionin a noisy information channel , Journal of Theoretical Biology 249 331–342. 2007

[16] Vladimir I. shCherbak,, Arithmetic inside the universal genetic code, BioSystems 70, 187–209, 2003

[17] Woese, C.R., Order in the genetic code. *Proc. Natl. Acad.* Sci., 54, 71–75. USA 1965

Design and implementation of a web-based medical record system (A case study of Federal Medical Centre, Lokoja, Nigeria)

Lizzy Oluwatoyin Ofusori
School of Management, IT and Governance
University of KwaZulu-Natal
Durban, South Africa
lizzyofusori@yahoo.co.uk

Umoru O. Jacob
Department of Computer Science
Federal College of Education, Okene
Kogi State, Nigeria
jacobumoru2004@yahoo.com

Bamidele Oluwade
Department of Computer Science
University of Ilorin
Kwara State, Nigeria
bamideleoluwade@computer.org

Adetoye Aribisala
Department of Mechatronics Engineering
Federal University Oye-Ekiti
Ekiti State, Nigeria
adetoye.aribisala@fuoye.edu.ng

***Abstract*— The maintenance, access and security problems associated with manual patient medical records have led to the search for alternative ways to enhance service delivery without compromising quality. This research study examines electronic medical records' potential to address these issues. The researchers explored how emerging integrated technology can reduce the cost of healthcare services using an Electronic Medical Record System. Alternative computer coding instructions such as Visual Basic 6.0 and Microsoft Access were used to build the database. Database unit and system testing was undertaken to check for data integrity. All the testing was successful and the application met all the required specifications.**

Keywords—Database, Electronic Medical records, Visual basic, Microsoft access.

I. INTRODUCTION

An electronic record system in a medical institution is a computer application developed to record information on patients registered in the hospital [1]. It suports efficiency and consistency in data storage and retrieval, as regards patients records in a healthcare system [2]. This comprehensive management system needs to be developed from scratch to fulfill a hospital's particular needs [3].

The term Electronic Medical Record (EMR) is used to describe an electronic version of the traditional paper record [3]. In some countries, it is described as an automated electronic medical record which includes all clinical data [4]. According to Torrey [5], the United States Federal Government has defined four basic functions that complete an Electronic Medical Record (EMR) system: computerized orders for tests, computerized orders for prescriptions, reporting of test results, and physician notes. However, Torrey [5] notes that this definition did not result in the desired standard. From the foregoing definition, one can conclude that an EMR also known as an Electronic Health Record (EHR) is an evolving concept which can better be described as an organized collection of electronic medical records on individual patients or a population. It is a digital format of a record that can be shared across various health care settings, or embedded in a network-connected enterprise-wide information system [6].

Emerson [7] highlighted, that, the two major advantages of e-medical records are that they save time and space. He notes that when a doctor has instant access to all a patient's information, such as laboratory tests, x-rays and information about allergies or prescriptions, he or she can take action immediately, thus saving time. Furthermore, while the Electronic Record System saves space, it does not render paper obsolete, but does considerably reduce the amount of paper used. Electronic medical records enable a health care team to coordinate and carefully manage records [5]. According to Lehmann, et al. [3] misunderstanding of information is minimized when it is typed and the system makes it easier for individual records to be located, downloaded or updated. Medicare and Medicaid Services [8] identified three advantages of EMRs, including potential to automate, and structuring and streamlining clinical workflow.

Studies have shown that computerized applications are part of medical information systems in most countries [1, 4, 6]. Health care authorities have adopted this system in order to promote efficient retrieval of patient records, statistics and research information [6].

An EMR is designed to accept data, process information, report, and use the information generated to improve the effectiveness and efficiency of medical services [2, 9]. It works on the same principle as a manual system, but the computer automates the process to make it fast and more accurate [3]

Manual paper-based systems to keep patient records are common in most hospitals and have existed for many years [7]. Studies have shown that the majority of doctors still find it hard to part with the manual system [2]. However, while it is easier for doctors to record medical data at the point of care, they require significantly more storage space than digital records [7]. It is stressful and time consuming for the healthcare provider to collate paper records stored in different locations [4]. An EMR system addresses these problems [5]. It also offers security features and data privacy and supports daily operations by eliminating duplicate data entry, thereby maintaining a detailed history of essential records [4, 6].

While the introduction of an EMR system can be a mammoth undertaking, the World Health Organization [10] is of the opinion that a computerized system has the potential to enhance the efficiency and effectiveness of patient record keeping. However, this requires that fundamental procedures are in place and that the system is well-organized.

According to Duchein [11], a major concern in relation to EMRs is that when such systems are introduced, doctors would experience a steep learning curve and, in essence, would have to become their own clerks, as they are required to enter a patient's symptoms or needs into the system. For her part, Torrey [5] expressed concern that intruders could hijack and exploit EMRs. Confidential medical information could be misused. Finally, Weeks [12] observed that the lack of a centralized database for all EMRs was a problem, as different health care centers and hospitals use different programs. This prevents all members of the medical team from having access to instant information, as some programs might not be compatible with others [12].

This research study was motivated by the fact that, as Lucas [13] observed, "African countries have a very low density health workforce, compounded by poor skill mix and inadequate investment". According to Weeks [12], the Nigerian government needs to provide basic communication infrastructure, stable electricity, and internet access to boost private initiatives to adopt information technology across the country. Furthermore, while many hospitals and health care centers are now adopting EMR systems, it remains uncertain how soon paper files will be entered into these systems [12]. Finally, it is important to ensure that the information and systems are uniform.

II. Systems Analysis

"Systems analysis is a problem solving technique that decomposes a system into its component pieces for the purpose of studying how well those component parts work and interact to accomplish the purpose" [14]. The current system of managing patients' records is analyzed in order to understand what is required (functional and non-functional) to develop an EMR system.

A. Shortcomings of the present system

The forms that are used to capture data are kept in files in a filing cabinet. Retrieval is both cumbersome and time consuming; moreover, this method occupies space and wastes resources. Inconsistency is the order of the day. Information contained in a particular cabinet may not agree with similar information in another cabinet as a result of records not being updated. This would not be the case if there was a centralized database.

B. System Requirements

The current system of managing patients and staff information was examined in order to understand the requirements (functional and non-functional) necessary for its operation. The researchers interacted with the staff of the Federal Medical Center, Lokoja. Document review and operational procedures were a major part of the enquiry.

- **Functional requirements:** Functional requirements are the factors that make the system perform its desired functions efficiently and effectively. They capture the intended behaviors of the system that can be expressed as the services, tasks or functions the system is required to perform. This is a systematic approach to building quality into the system.
- **Non-functional requirements:** These have to do with the qualitative requirements that are not necessarily for the system to perform the desired functions. They include:
  i. consistency in design and use of colors
  ii. simplicity in design and user friendliness,
  iii. the system is secure and fully fraud proof

III. System Design

Whitten, et al. [15] defined system design as the tasks that focus on the specification of a detailed computer-based solution. Having analyzed the current manual system employed by the Federal Medical Center Lokoja to manage patient and staff information, the researchers set out to

design a more robust system that would address the shortcomings of the current system.

A. Database design

The database design was implemented using Microsoft Access 2003. This is a relational database management system that organizes data in tables that are related or linked to one another. Each table consists of rows and columns called fields.

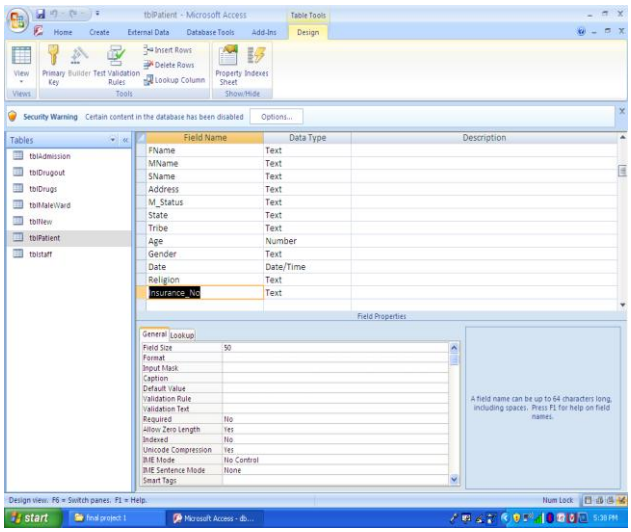


Fig 1: Database table structure design

Main Menu: The menu editor was used to create a standard bar which contains the main menu items based on categories; with each category containing sub menu items. This is shown below.

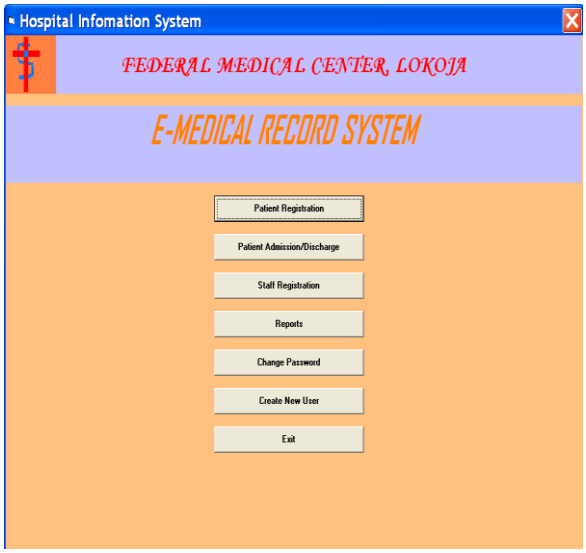


Fig 2: Main menu

Patient Registration: This form is used to enter all the necessary information about the patient. The interface is shown below:

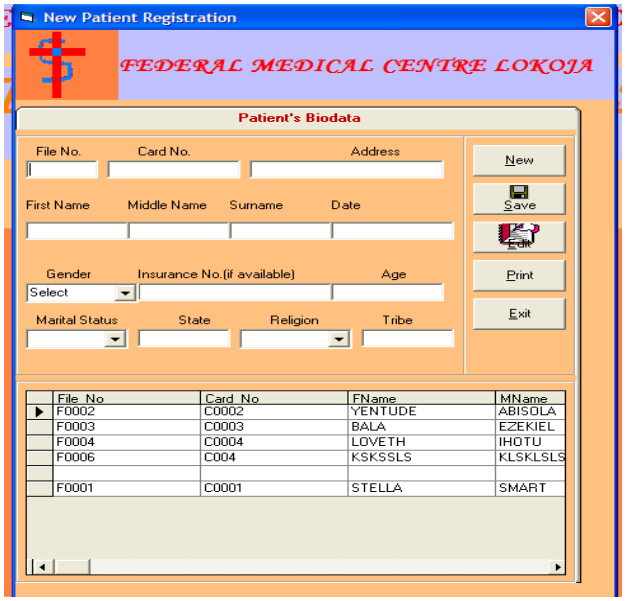


Fig 3: Patient registration

Patient Admission/Discharge module: This form records when a patient is admitted or discharged and the treatment prescribed.

Fig 4: Patient Admission/Discharge module

Common Features

The common features of almost all the forms are:

- **Save:** After all the required entries for a record have been made, a click on this button saves the record in the database.
- **Add new:** used to add a new record to the database; a click on this button clears the form for a new entry.
- **Delete:** used to remove a particular record from the database.
- **Edit:** edit is used to retrieve a particular record from the database for the purpose of altering text fields.
- **Update:** After a record has been retrieved or corrected, this button is used to effect changes to the same record in the database. This does not create a new record.
- **Exit:** This button closes the active displayed form.

IV. LINKING VISUAL BASIC WITH MS ACCESS

Visual Basic (VB) has the capability to access and manipulate data from an external database stored in a different package such as Microsoft Access. VB directly supports database files in several formats. The native format is Microsoft Access, using the jet database engine. Linking a VB application to a Microsoft Access database requires a Dynamic Link Library (DLL) called the Microsoft ActiveX Data Object Library and a provider called the Microsoft Jet 4.0 OLE DB provider. Dynamic link libraries are system files with .dll

- **Database Implementation**

This is the collection of related records (information) on patient information, staff information, and treatment details stored elsewhere for easy retrieval and exploration. The VB is linked to the database and the database is implemented using MS-Access.

MS-Access is a database application that can create database files using relational models. Tables can be created and the data in the tables can be stored and manipulated as required. Relationships can also be established among tables with a common field.

- **Security**

Access to the database is restricted and controlled using a password. Unauthorized access to the system is thus prevented. Before a user logs on to the system, he or she is required to enter his/her password for validation.

- **System Testing, Debugging and Compilations**

This verifies that the whole system works according to the design. Each unit module in the software design is tested in order to ensure that there is no error as a result of an incorrect code, syntax error and typographical error. After successfully testing the unit modules, the same testing is done for the system as a whole in order to ensure that the program is working as required. The unit and system module are brought together after testing in order to integrate the program into a single module.

During the Debugging stage, the integrity of the program was tested by comparing program performance to required specifications. The level of user friendliness was also evaluated by observing how easy it is for users to use the system.

The final step in the implementation of this system was the compilation of the project. The purpose was to compile the code so that it becomes an executable file. This prevents access to the codes and makes the application standard software that can be installed and run on any system that meets the minimum specifications. It also makes the application independent of the existence of VB and MS Access that were used to develop it.

V. RECOMMENDATIONS

The researchers recommend that management of Federal Medical Centre Lokoja considers the adoption of this software. If possible, an expert in software development should be employed to ensure that the software is of high utility. To achieve the goals of this project, the following are also recommended.

i. Staff Training: Personnel that will be in charge of the new system should be given adequate training on its use.
ii. Changeover Method: The researchers suggest the use of a parallel changeover method, which involves running the old and the new systems simultaneously for a "test period" during which the performance of the new system is carefully observed. This would also enable management and personnel that will be using the new system to get used to the system before the old system is discontinued.
iii. The office should be marked "Out of bounds to unauthorized personnel".

VI. CONCLUSION

The Electronic medical system was developed using Visual Basic 6.0 programming language. This initiative is likely to reduce service inefficiencies and facilitate the retrieval of clients' health records, thereby reducing turnaround time from the time a client is attended to by a health official to their treatment based on their medical assessment. The use of such technologies would not only improve the client's service experience but enhance citizens' confidence in the integrity of the healthcare system.

REFERENCES

[1] R. Hillestad, J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, *et al.*, "Can electronic medical record systems transform health care? Potential health benefits, savings, and costs," *Health Affairs*, vol. 24, pp. 1103-1117, 2005.

[2] J. King, V. Patel, E. W. Jamoom, and M. F. Furukawa, "Clinical benefits of electronic health record use: national findings," *Health services research*, vol. 49, pp. 392-404, 2014.

[3] C. U. Lehmann, K. G. O'Connor, V. A. Shorte, and T. D. Johnson, "Use of electronic health record systems by office-based pediatricians," *Pediatrics*, pp. peds. 2014-1115, 2014.

[4] M. Amatayakul, *Electronic health records: A practical guide for professionals and organizations*: Amer Health Information Management, 2007.

[5] T. Torrey, "The Benefits of Electronic Medical Records (EMRs)," ed: ABOUT. COM, Feb, 2008.

[6] A. Appari, M. Eric Johnson, and D. L. Anthony, "Meaningful Use of Electronic Health Record Systems and Process Quality of Care: Evidence from a Panel Data Analysis of US Acute-Care Hospitals," *Health services research*, vol. 48, pp. 354-375, 2013.

[7] P. Emerson, "What is the records management? How to Manage Your Records? A Guide to Effective Practices," *Chambridge: ICSA Pub*, 1989.

[8] C. f. Medicare and H. Medicaid Services "Medicare and Medicaid programs; electronic health record incentive program. Final rule," *Federal register*, vol. 75, p. 44313, 2010.

[9] A. Likourezos, D. B. Chalfin, D. G. Murphy, B. Sommer, K. Darcy, and S. J. Davidson, "Physician and nurse satisfaction with an electronic medical record system," *The Journal of emergency medicine*, vol. 27, pp. 419-424, 2004.

[10] World Health Organization, *International statistical classification of diseases and related health problems* vol. 1: World Health Organization, 2004.

[11] M. Duchein, "Obstacles to the Access, Use and Transfer of Information from Archives: A RAMP Study," 1983.

[12] R. V. Weeks, "Electronic health records: managing the transformation from a paper-based to an electronic system," 2013.

[13] A. O. Lucas, "Human resources for health in Africa," *British Medical Journal*, vol. 7524, p. 1037, 2005.

[14] M. J. Neubauer, *A Systems Analysis of Information Technology and the Use of WLANs Implemented by an FBI Field Office for Crisis Response Incidents: The Columbia Field Office Case Study*: ProQuest, 2007.

[15] L. Whitten, V. M. Barlow, and L. Bentley, *Systems analysis and design methods*: McGraw-Hill Professional, 1997.

Proceedings of the 1St International Conference of IEEE Nigeria Computer Chapter In collaboration with Department of Computer Science, University of Ilorin, Ilorin, Nigeria - 2016

Application of Smith-Wateman and Needleman-Wunsch Algorithm in Pairwise Sequence Alignment of Deoxyribonucleic Acid

Bello Hassan Kehinde
Department of Computer Science,
Federal Polytechnic, Offa
Kwara state, Nigeria
hassan.bello@fedpoffaonline.edu.ng

Gbolagade Kazeem Alagbe
Department of Computer Science
Kwara state University, Malete
Kwara State, Nigeria
kazeem.gbolagade@kwasu.edu.ng

***Abstract*— in bioinformatics, sequence alignment is solely concerned with the arrangement of Deoxyribonucleic Acid (DNA), Ribonucleic Acid (RNA) and Proteins sequences in order to identify region of similarity that may be consequence of functional, structural or evolutionary relationship between the sequences. DNA is the primary genetic and hereditary material which contains instructions that direct the cell in all living organisms. It is made up of chemical building block called nucleotides which contains nitrogen bases adenine, cytosine, guanine and thymine abbreviated as A,C,G and T respectively. Genetic database hold a large amount of raw data, human genome alone has nearly 3 billion DNA base pairs. To search through this data and infer a meaningful relationship from it, molecular biologists rely on efficient and effective computer algorithm. This paper introduces Smith-waterman algorithm and Needleman-Wunsch algorithm (that use dynamic programming) and we present solution that performs the alignment for any given pair of sequence of molecules. The alignment value is obtained by use of Smith-waterman and Needleman-Wunsch algorithms.**

Keywords*— Sequence alignment, Deoxyribonucleic Acid, Ribonucleic Acid, Protein, Database, Nucleotide.

I. INTRODUCTION

In bioinformatics, searching for Deoxyribonucleic Acid (DNA), Ribonucleic Acid (RNA) or Protein molecules involve several methods and algorithms. The nucleotides in DNA molecules are arranged in several ways. The order of arrangement of these nucleotides determines the biological instruction in a strand of DNA.

As evolutionary relationships believe that certain numbers of amino acid residues in protein sequence are conserved, the easiest way to assess the relationship between two sequences will be by counting the number of identical and similar amino acids. This is done by sequence alignment.

The rest of the article is organized as follows: Section II explains the background, section III discuss sequence alignment, section IV type of sequence alignment, section V methods of sequence alignment, section VI discuss Neddlema-wunsch and Smith-waterman algorithms and finally in section VII we present the conclusion of the article.

II. BACKGROUND

The era of modern sequence alignment techniques began in 1970 with the publication by Needleman and Wunsch of a dynamic programming method which could be used to make a global pairwise alignment of two protein sequences [1]. Soon after, Seller (1974) improved over the algorithms [2] which is about to fulfill the request of biology by measuring the metric distance between sequence [3].

In 1981, Smith and Waterman made a modification on Needleman-Wunsch publication to create the local alignment algorithm known as the Smith–Waterman algorithm [3]. Both the Needleman–Wunsch and Smith–Waterman methods belong to a class of algorithms called dynamic programming algorithms [4]. This class of algorithms can find optimal solutions to problems but can take a long time to run, especially in complicated cases or for large data sets. These two algorithms are the most accurate pairwise alignment algorithms in existence. Nearly all of the newer local pairwise alignment algorithms use a two step approach to reduce the running time. The first stage uses heuristics to search for areas which have a high probability of producing alignments. Next, these areas are passed to a dynamic programming algorithm such as the Smith–Waterman algorithm for true alignment. The most commonly used two step approaches are FASTP/FASTA, the BLAST family of algorithms [4].

III. SEQUENCE ALIGNMENT

Sequence alignment in bioinformatics compute similarities between two sequences [5]. It describes the arrangement of DNA or RNA or Protein sequences as to identify the regions of similarity that may be a consequence of functional, structural or evolutionary relationship between the sequences. Alignment finds level of similarity between query sequence and different database sequence. This can be accomplished by

using dynamic programming approach. Sequence alignment can be used to predict the function and structure of a new function.

A. *Why Sequence Alignment?*

Sequence alignment is used to perform similarity check and we perform a similarity search to learn if our sequenced DNA can be found in a public nucleotide database (i.e. it has already been cloned by others) and/or whether it is evolutionally related (i.e. homologous) to other sequences. In a simple similarity search, one can compare a sequence with sequences found in an entire nucleotide database. This information is important because we will be able to identify the precise location and sequence of human genes will allow us to better understand genetic diseases.

B. *Public Databases*

Primary Databases International Nucleotide Sequence Database (INSD) consists of the following databases.

- DNA Data Bank of Japan (National Institute of Genetics) [6], [7]
- EMBL (European Bioinformatics Institute)
- GenBank (National Center for Biotechnology Information)

The three databases, DDBJ (Japan), GenBank (USA) and European Nucleotide Archive (Europe), are repositories for nucleotide sequence data from all organisms. All the three databases accept nucleotide sequence submissions, and then exchange new and updated data on a daily basis to achieve optimal synchronization between them. These three databases are primary databases, as they house original sequence data.

IV. TYPE OF SEQUENCE ALIGNMENT

A. *Pairwise sequence alignment*

This is the alignment between two sequences in order to find the sequences that are similar. The input sequence also called query sequence is aligned with each target sequence in the database. Each alignment between query sequence and target sequence is one pairwise alignment. For each pairwise alignment, a score is allocated as to indicate level of similarity between query and the corresponding target sequence.

B. *Multiple sequence alignment*

In multiple sequence alignment, the objective is to find a common alignment for multiple sequences. Several approaches for multiple sequence alignment have been proposed. Initial implementations were based on an extension of the Smith Waterman algorithm to multiple sequences. This implementation, which is based on dynamic programming, generates an optimal solution, but is computationally very intensive. More recent approaches incrementally build multiple sequence alignment by using heuristics. [8]

V. METHODS OF SEQUENCE ALIGNMENT

A. *Heuristic Algorithm*

The algorithm quickly finds solutions close to the best among all possible ones, but accuracy is not guaranteed (they may be considered as approximately and not accurate algorithm). Example FASTA *and BLAST*

1) The FASTA (FAST –All) Algorithm

The FASTA [9] was developed by William R Pearson and David J Lipman in 1985 **[10].** *The algorithm can be used to search similarity sequences in DNA or Proteins sequence within time. FASTN (Fast Nucleotides) is used for the analysis of Nucleotide and FASTP (Fast Proteins) is used for the analysis of Proteins. The combined package known as FASTA (Fast-All).*

2) The BLAST (Basic Local Alignment Search Tool)

Basic Local Alignment Search Tool [11], [12] algorithm and program were designed by Stephen Altschul, Warren Gish, Webb Miller, Eugene Myers, and David J. Lipman [13]. It finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. It originated in 1990 as an improvement to FASTA.

BLAST algorithm can answer the following questions:

- where does a certain DNA sequence originate
- which DNA sequence share similarity with the sequence in question

B. *Dynamic Programming*

Dynamic programming (DP) tends to break a complex problem into sub-problems and finds solution to the sub-problems then come up with the solution to the complex problem. DP is simply the method "*Divide and Conquer*". Typical examples of algorithms that use DP technique are

- Global Pairwise alignment Algorithm (Needleman Wunch)
- Local Pairwise alignment Algorithm (Smith-Waterman algorithm)

1) Global Pairwise Alignment

A global pairwise alignment is a sequence alignment over the entire length of two or more nucleic acid or protein sequences. In this method, alignment is between the complete sequence A and complete sequence B. Implemented by Needleman-Wunsch Algorithm. An algorithm developed by

Saul B. Needleman and Christian D. Wunsch in 1970 [1]. It was one of the first applications of *Dynamic* programming to compare biological sequences.

Example of Global Pairwise Alignment

```
ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA
|||||||||| |||     |||||||  ||||||||||||||| |||||
'ACTACTAGATT----  ACGGATC--  GTACTTTAGAGGCTAGCAACCA
```

2) Local Pairwise Alignment

A Pairwise Local alignment is an alignment of two sub-regions of a pair of sequences. Alignment between a sub-sequence of A and a sub-sequence of B.

It identifies the most similar sub-region shared between two sequences. Implemented by Smith- Waterman algorithm. This is one of the most advanced and sensitive pairwise sequence comparison algorithms available [14]. In the arrangement of sequence, we may see that letters are matched, mismatched, Inserted or Deleted (InDel).

Matched: If the two letters in the sequences are the same
Mismatched: if the two letters in the sequences are different
Indel (INsetion or DELetion): A letter aligns to a gap in the other sequence.

Example of Local Pairwise Alignment

```
ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA
||||  |||||||  ||||||||||||||||  |||||||
TACTCACGGATGAGGTACTTTAGAGGC
```

The two methods of alignments mentioned above are mostly defined by *Dynamic Programming* approach for aligning two different sequences.

VI. NEEDLEMAN-WUNSCH AND SMITH-WATERMAN ALGORITHM

The steps in Needleman-Wunsch and Smith-Waterman Algorithm

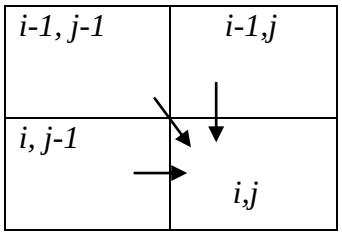
- Initialization of matrix
- Matrix filling with the appropriate scores
- Trace back the sequences for a suitable alignment [6].

Various algorithms are used in computing alignment of nucleotides. Needleman and Wunsch(1970) were the first to introduce heuristic algorithm for alignment which is used to calculate homology between sequences. The algorithm is based on calculation of global alignment [1]. Soon after Seller (1974) improved over the algorithms [2] which is about to fulfill the request of biology by measuring the metric distance between sequence [3].

A. The Smith-Waterman Algorithm

The algorithm was first proposed by Temple F. Smith and Michael S. Waterman in 1981 [3]. Like the Needleman–Wunsch algorithm, of which it is a variation. The Smith–Waterman Algorithm (SWA) performs local sequence alignment; that is, for determining similar regions between two strings or nucleotide or protein sequences [16]. Instead of looking at the total sequence, the Smith–Waterman algorithm compares segments of all possible lengths and optimizes the similarity measure. The algorithm is based on calculation of local alignment and also permits deletion and insertion of arbitrary length. It is used to implement dynamic programming technique. When searching for database for sequence homology, Smith-Waterman algorithm is the most accurate but time consuming.

$$X(i,j) = Max \begin{cases} 0 \\ X(i-1,j-1) + S(a_i,b_j) \text{ match/mismatch} \\ X(i-1,j) + g \\ X(i,j-1) + g \end{cases}$$



Example

Consider the sequence B1: A C A C A C T A
B2: A G C A C A C A

Scoring parameter: match = +2, mismatch = -1 and gap penalty = -1

To compute the alignment, we have to follow the 3 steps mentioned earlier.

Step 1: matrix initialization

△	-		C	A	C	A	C	T	A
-	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Step 2: filling the matrix

Use the general formula of SWA below and apply the scoring parameters to fill the table.

△	-	A	C	A	C	A	C	T	A
-	0	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0	2
G	0	1	1	1	1	1	1	0	1
C	0	0	3	2	3	2	3	2	1
A	0	2	2	5	4	5	4	3	4
C	0	1	4	4	7	6	7	6	5
A	0	2	3	6	6	9	8	7	8
C	0	1	4	5	8	8	11	10	9
A	0	2	3	6	7	10	10	10	12

Scoring matrix

Step 3: The Trace back
How to trace back:
Start with H(i,j) than is maximal
Follow directions (maximal entries in DP)
Stop when you reach H(i,j) = 0

△	-	A	C	A	C	A	C	T	A
-	0	0	0	0	0	0	0	0	0
A	0								
G	0								
C	0								
A	0								
C	0								
A	0								
C	0								
A	0								

Best value = cell (8,8) with $H_{8,8} = 12$
This can be cross-check by associated alignment:

Sequence1: A – C A C A C T A
Sequence2: A G C A C A C – A
Number of matched = 7
Number of mismatched =2
Best value = 7(2) + 2(-1) = 12

Fig. 1 and fig. 2 show how best value can be obtained through trace back and associated alignment.

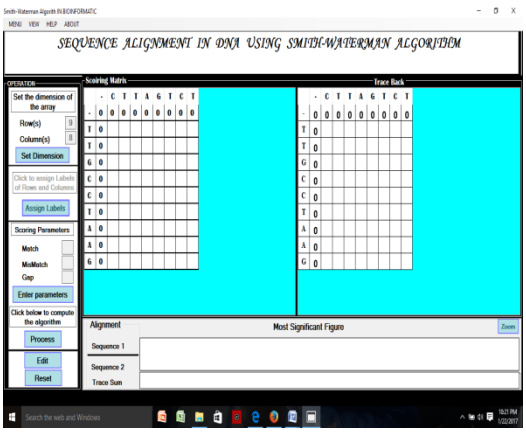


Fig. 1. (initialization of scoring values)

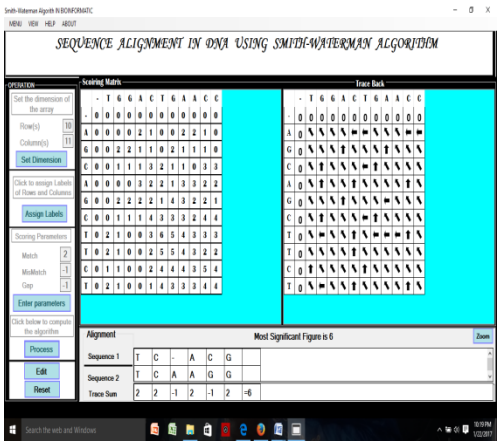


Fig. 2. (trace back and best value)

B. The Needleman-Wunsch algorithm

To use Needleman-Wunsch algorithm, a matrix is created for the purpose of comparing the query and target sequences

$$X(i,j) = Max \begin{cases} X(i-1,j-1) + S(a_i,b_j)\ match/mismatch \\ X(i-1,j) + g \\ X(i,j-1) + g \end{cases}$$

$X(i,0) = 0, \quad X(0,j) = 0$

Where,

m is the length of a, n is the length of b

X(i,0) = 0, X(0,j) = 0

X(i,j) => maximum similarity score between the two sequences

$s(a_i,b_j)$ => the similarity score of comparing sequence Ai to sequence Bi

g is the gap alignment and i,j describes rows and column.

To use Needleman Wunsch algorithm on the following sequences,

Sequence1: CTTCA and

Sequence2: CTACA

Scoring: Match =+5, Mismatch = -1 and gap = -2

		C	T	T	C	A
	0	-2	-4	-6	-8	-10
C	-2	5	3	1	-1	-3
T	-4	3	10	8	-6	4
A	-6	1	8	9	7	11
C	-8	-1	6	7	14	12
A	-10	-3	4	5	12	19

Best value = cell (5,5) with $H_{5,5} = 19$

This can be cross-check by associated alignment:

S1: C T T C A

S2: C T A C A

Number of matched = 4

Number of mismatched = 1

Best value = 4(5) + 1(-1) = 19

Smith-Waterman algorithm can also be used to compute the best value of b the following two sequences:

Sequence1: CTTCA and Sequence2: CTACA with Scoring: Match =+5, Mismatch = -1 and gap = -2

		C	T	T	C	A
	0	0	0	0	0	0
C	0	5	3	1	5	3
T	0	3	10	8	6	4
A	0	1	8	9	7	11
C	0	5	6	7	4	12
A	0	3	4	5	2	19

Best value = cell (5,5) with $H_{5,5} = 19$

This can be cross-check by associated alignment:

S1: A C A T C

S2: A C T T C

Number of matched = 4

Number of mismatched = 1

Best value = 4(5) + 1(-1) = 19

It produced the same value as obtained in Needleman-Wunsch algorithm.

VII. CONCLUSION

In bioinformatics, Needleman-Wunsch and Smith-Waterman Algorithms applied dynamic programming to provide solution in the alignment of nucleotides. The matrix of Needleman-Wunsch may contain negative numbers whereas the highest value in SWA matrix is zero. In the process of alignment of sequence of DNA or proteins, short sequences can be aligned /adjusted manually but in case of very high number of sequences which is not humanly possible, an algorithm may be designed to solve or compute the alignment. The Needleman-Wunsch is appropriate for finding the best alignment of two sequences with similar length and searches across the entire lengths. It has been demonstrated that global alignment can also be used to find the similarity between subsequences of two sequences if the subsequences are of the same length.

REFERENCES

[1] Needleman S.B and Wunsch C.D (1970) " A general method applicable to the search for similarities in the amino acid sequence of two proteins " Journal of Molecular biology 1970, 443-453.

[2] Sellers P.H. (1974) " Journal of applied Mathematics (Sion), 26, 787-793.

[3] Smith Temple F., and Waterman Michael (1981) "identification of Common Molecular Subsequences" (PDF). Journal of Molecular Biology 147: 195–197. doi:10.1016/0022-2836(81)90087-5. PMID 7265238

[4] D. Edwards et al. (eds.) (2009), "Bioinformatics: Tools and Applications", DOI 10.1007/978-0-387-92738-1_2, © Springer Science Business Media, LLC 2009

[5] Robert S. Harris (2007) "Improved Pairwise Alignment Of Genomic DNA" A Thesis in Computer Science and Engineering, College of Engineering Pennsylvania State University

[6] Tateno Y, Imanishi T, Miyazaki S, Fukami-Kobayashi K, Saitou N, Sugawara H, et al. (2002). "DNA Data Bank of Japan (DDBJ) for genome scale research in life science.". Nucleic Acids Res. 30 (1): 27–30. doi:10.1093/nar/30.1.27. PMC 99140. PMID 11752245.

[7] Jump up Kaminuma E, Kosuge T, Kodama Y, et al. (January 2011). "DDBJ progress report". Nucleic Acids Res. 39 (Database issue): D22–7. doi:10.1093/nar/gkq1041. PMC 3013661.PMID 21062814.

[8] Vipin Chaudhary, Feng Liu, Vijay Matta, Xiandong Meng, Anil Nambiar, Ganesh Yadav, Laurence T. Yang "Parallel Implementations of Local Sequence Alignment: Hardware and Software" St. Francis Xavier University, Parallel & Distributed Computing Laboratory,

Antigonish, NS, B2G 2W5, Canada. Wayne State University, Detroit, MI, USA.

[9] W. R. Pearson and D. J. Lipman, "Improved tools for biological sequence comparison.," Proceedings of the National Academy of Sciences of the United States of America, vol. 85, no. 8, pp. 2444–2448, 1988. View at Publisher · View at Google Scholar · View at Scopus

[10] Lipman, DJ; Pearson, WR (1985). "Rapid and sensitive protein similarity searches". Science 227 (4693): 1435–41. doi:10.1126/science.2983426. PMID 2983426.

[11] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," Journal of Molecular Biology, vol. 215, no. 3, pp. 403–410, 1990. View at Publisher · View at Google Scholar · View at Scopus

[12] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," Nucleic Acids Research, vol. 25, no. 17, pp. 3389–3402, 1997. View at Publisher · View at Google Scholar · View at Scopus

[13] Altschul, Stephen; Gish, Warren; Miller, Webb; Myers, Eugene; Lipman, David (1990)." Basic Local Alignment Search tools". Journal of Molecular Biology 215 (3): 403–410. doi:10.1016/S0022-2836(05)80360-2. . PMID 2231712.

[14] Tim H., Jacob V, Jack A.M Peter M.G (2006) "testing statistical significance scores of sequence comparison methods with structure similarity" BMC bioinformatics 20067:444 doi 10ii86/1471-2105-7-444

[15] Kwame O. and Edward Y.B (20012) "A smith-Waterman algorithm accelerator based on RNS" IJECE ISBN1974-2166 VOL5 pp 99-112

[16] Gotoh O. (1982) " An improved algorithm for matching biological sequences; Journal of Molecular Biology, 162(3):705-708.

Proceedings of the 1St International Conference of IEEE Nigeria Computer Chapter In collaboration with Department of Computer Science, University of Ilorin, Ilorin, Nigeria - 2016

A survey of literature on paternity test software

Sharon Ikechi
Computer Science Department
University of Ilorin
Ilorin, Nigeria
Sharon.kechi@gmail.com

Bamidele Oluwade
Computer Science Department
University of Ilorin
Ilorin, Nigeria
deleoluwade@yahoo.com

***Abstract*— PedExpert, Genolab and EasyDNA are three paternity test software for calculating the probability of likelihood ratio for paternity tests. The paternity and kinship test software are surveyed based on the concept of the application of Bayesian network in complex paternity cases such as when the DNA profile of the alleged father is not known. Other concepts include Bayes theorem for calculating paternity index, cloud computing for high computing speed and storage capacity, genetic mutation in paternity testing, user friendliness and pedigree design. This paper explains the concept of paternity testing and gives descriptions of the features of the three paternity test software.**

Keywords— ***Paternity testing; Kinship analysis; DNA profiling; computer software; Genetic mutation; Pedigree.***

I. INTRODUCTION

Paternity is whether an alleged man is biologically the father of a child or not. Test carried out to determine paternity is called a paternity test. A maternity test is used to determine the biological mother of a child. The technique for carrying out a paternity test is the same for maternity test. The concept of paternity testing in scientific journal can be originally traced to 1956 [1]. In 1996, the United States recorded 420,740 "legal" paternity tests. The need for paternity test arises legally, in paternity courts, socially, in private paternity disputes, or medically, in prenatal diagnosis, genetic counseling and transplantation [2]. The concept of complex paternity case arises in paternity cases where the DNA information of a person involved in the test is not known due to the fact that the person may be dead or missing.

Paternity test software has incorporated methods for determining paternity in complex and noncomplex cases. There are various methods or techniques for carrying out paternity test. These methods have been improved over time with the recent method giving high functionalities. The oldest method is the ABO blood group typing. The ABO group was discovered in 1901 and the Rhesus group in 1937. In humans, ABO and Rhesus factor (Rh factor) are two important classifications to describe blood types [3]. ABO means blood group A, B and O. Rhesus factor could be positive or negative. ABO blood group typing is a method of carrying out paternity test by determining the blood group type of the father and child and checking for compatibility. This method could not produce absolute proof of fatherhood.

The second method involved HLA Analysis which emerged in the 1970s. HLA means Human Leukocyte Antigen. HLA is a super gene that contains the major proteins in humans and it is found in the white blood cells. It was a better technique because unlike the ABO blood group typing that had only four possible blood groups, this had about hundreds of proteins contained in the HLA that could be used for analysis. This method could rule out the fact that a man was not the father but it could not prove that the man was the biological father.

In the 1980s DNA testing method was discovered and it had 99.99 percentage accuracy when the man is the biological father of the child. The techniques used in DNA testing are polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP). DNA means Deoxyribonucleic acid and it is the basic genetic material contained in all living most cells of the body. DNA is every person's genetic "blueprint" and everyone has a unique DNA pattern. Half of their DNA is inherited from their mother and half from their father and these DNA materials are known as alleles. When the child's DNA is compared to an alleged parent and no match exists, that person is excluded as the biological parent and a probability of paternity less than 10% is gotten. If there is a match in the DNA patterns, a probability of 90% or greater is calculated thus establishing a biological relationship [4].

The techniques used in DNA testing are polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP). Just from the word chain reaction, a small cause leading to many effect, the polymerase chain reaction is a technique for enlarging a short sequence of DNA to generate million sequences of that particular DNA sequence. It was developed in 1983. The short DNA sequences are called primers. RFLP involves the use of enzymes to break down a sample of DNA into fragments

and separating them according to their lengths. This technique was the first to be used in generating DNA profiles which are used in paternity tests. Modern genetics involves not only blood samples for DNA testing but also hair follicle sample, buccal swabs, saliva and semen sample. DNA paternity tests are based on the fact that each individual has two alleles. One allele is inherited from the biological father and the other from the biological mother, except in monozygotic twins [5].

DNA profiles are a set of DNA sequences that are peculiar from individual to individual. DNA profiling is a technique for obtaining the DNA profiles of an individual. It was developed in 1988 by Sir Alec Jeffreys. PCR and RFLP are the technologies used in DNA profiling and they have opened a new era of forensic DNA profiling. DNA profiling works by targeting short tandem repeats on different chromosomes and uses the PCR technique to amplify them. Short Tandem Repeats (STR) are the regions in a DNA sequence where nucleotides are repeated a number of times. The process of capturing and recording STR along with their specific location on the chromosome is known as DNA profiling. The STR will be useful in comparing the sequences of the father and child for matches, though they are inherited but they vary. Each individual has a unique DNA profile because several STR markers are used for the sequencing which makes it perfect for a paternity test.

This paper presents an overview on three paternity test software which are PedExpert, Genolab and EasyDNA. It includes their platform, advantages, basic functionality and interface.

II. PATERNITY TEST

A paternity test is a computational procedure that is carried out by a paternity test software. A Paternity test software is a computer program that can calculate the probability of paternity during a paternity test. They are used in DNA Laboratories in aligning the DNA sequences in the DNA profiles and calculating the percentage of a match or a mismatch. These programs have specific instructions known as algorithms. The results gotten from paternity test software is called a Paternity Index (PI). The software uses statistical inference to proof paternity. This can be achieved by calculating a ratio of the likelihood of obtaining the observed set of findings given that the alleged father is the true one (X) over the same likelihood in the hypothesis of a random man being the father (Y) [6].

Paternity test is carried out based on statistical inference using Bayes theorem. It uses probability to measure the degree of certainty of the alleged man being the father of the child or not, by giving a prior probability and calculating a posterior probability. Paternity test result could be wrong due to factors such as human error, genetic mutation, related dads and so on. The factor that can be controlled by the software is genetic mutation. Genetic mutation occurs when there are changes in the DNA. However, this can only occur when the change is reflected in the same Loci of the DNA structure of the alleged father and child. To prevent error in paternity testing, the software has to incorporate a method of handling genetic mutations known as Bayesian network.

Paternity test can be categorized based on the DNA information available, how willing the alleged father is to be tested and the reason of the test. [7]. Three types of paternity tests according to Maureen Young are:

1. Informational Paternity Test: This test is carried out just for the information sake which is to know the biological father of the child. This type of test cannot be used for legal or organizational purposes.

2. Legally-Admissible Paternity Test: This type of paternity test is for legal purposes and it can be presented in a court room during a case. After the paternity test is conducted. A third party is required to collect the test results from the laboratory to prevent any form of fraud.

3. Family Relationship Test: This type of paternity test is carried out to find the biological father of a child in a case where the father is missing or deceased.

III. PATERNITY SOFTWARE OVERVIEW

The survey conducted in this research paper includes an analysis of three paternity test software. They are Ped expert, Genolab, Easy DNA and CERVUS. These software are used in a paternity test and they will be analyzed based on the following criteria: platform, advantages and interface.

1. Ped Expert: Ped Expert is a computer program that applies Bayesian Networks to human paternity testing. It is used in complex situations such as when genetic information of the alleged father which is required is missing either because he is dead or missing. Bayesian networks tries to reconstitute his DNA profiles from the man's relatives. Its algorithm converts the DNA information into Bayesian Networks. One important feature of Ped Expert is that it provides a platform for the identification of genetic mutation information. It is great for complex situations due to its user friendliness and time saving capability.

Advantages
1. It involves the creation of the structure of Bayesian Networks from the family
pedigrees.
2. Genetic information of individuals in the pedigree structure will be contained in the
Bayesian Networks.

3. It uses the Mendelian principle to construct tables of conditional probabilities which are contained in the Bayesian Networks.
4. Its algorithm converts pedigree structure and DNA information into Bayesian
Networks that is run on the SMILE environment.

Interface

Interface of PedExpert software is given in the figure below [6].

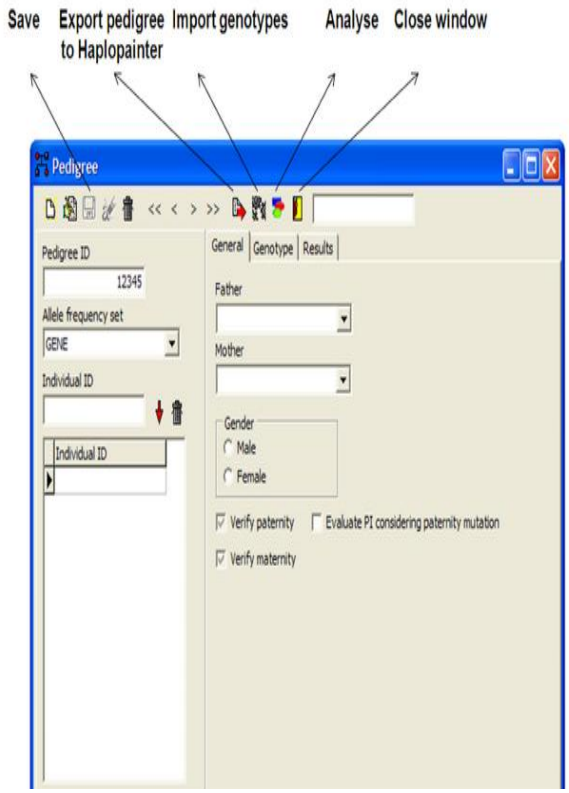


Fig. 1. The interface of PedExpert software. The allele number inherited from the father and mother by the child is supplied as input to the two tabs titled father and mother respectively. Paternity test and maternity test can be carried out. Paternity index can also be calculated.

1. Save : This icon is used to save the task performed by the user to a file on the system.
2. Import Genotypes : This icon is used to input the DNA information into PedExpert by
Importing the DNA information from a Microsoft Excel file.
3. Export Pedigree to Haplopainter: This icon is used to automatically export the data to the Haplopainter. Haplopainter is a pedigree drawing software which ensures that all the family members have been constructed in the pedigree structure.
4. Analyse: This icon calculates the paternity index or the likelihood ratio of the information located within the network. The results tab calculates the combined paternity Index using Bayes theorem and a priori probability of 0.5 and the results can be exported as a report to a Microsoft Excel spreadsheet.
5. Close Window: This icon exits the user from the PedExpert software [6].

Platform
Ped Expert works on a Microsoft Windows Based platform and its implementation
platform is Microsoft.NET using Borland Delphi 2006.NET as its development environment. Ped Expert operates on the SMILE environment and it makes use of SMILE.NET which is a version of SMILE within a Dynamic link library. SMILE.NET contains the main classes and methods of the SMILEAPI [6].

2. Genolab: Genolab is a software that adopts the cloud technology in paternity testing. It was especially developed to face the future challenge in forensic data analysis and biostatistics in relation of data amount and analysis time by the usage of cloud technology and development of algorithms. [8].
The aim of developing Genolab was to be able to carry out paternity tests of complex cases in less time by using cloud computing technologies to provide computer power and storage capacity. The basic functions of Genolab are paternity testing, kinship analysis and DNA evidence interpretation.

Platform
A user can connect to the Genolab software by creating an account and logging in
through a personal computer, tablet or smart phone. It has a security feature that helps to protect its users from unauthorized access. Its security mechanisms include encryption of data and SSL-encrypted connections.

Advantages
1. All algorithms used in Genolab are developed and optimized for cloud technology.
2. It includes a huge database that contains population data that can be used by users to perform calculations.
3. It secures the users data and operations by encrypting the data over the connections.
4. The program can handle potential mutations according to different models like step model, equal or decreasing probability with increasing step size as well as subpopulations [9].

Interface

Interface for Genolab software is given in the figure below [8].

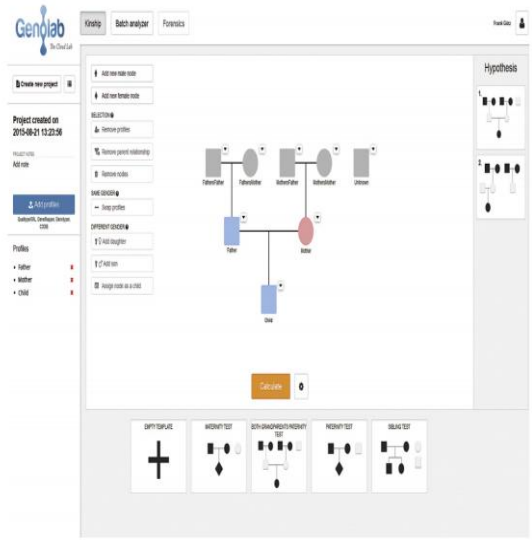


Fig. 2. The interface of Genolab software. The pedigree design shows the relationships that exist among the persons to be tested

1. Add Profile: This icon allows the user to upload profile data of various persons. The file format for the profile data are Qualitype XML, CODIS, GeneMapper CSV and Genotyper CSV.
2. Graphical pedigree designer: it is a tool for creating pedigrees. It also has templates of ready made pedigrees for fast implementation. The DNA profile of each person can be added one after the other to the pedigree.
3. Calculate: The calculate tab directly executes calculations to obtain the likelihood ratios and presents reports that show the likelihood ratio and probabilities of the markers of the DNA profiles added to the pedigree.
4. Batch Analyzer: This allows hundreds of kinship analysis to be performed at once.

Genolab can handle a huge number of profiles which could be shown by testing with up to 100.000 families and 100.000 missing persons in one analysis run [8].

3. Easy DNA : Easy DNA is a software package that consists of four applications for kinship and paternity determinations and statistical calculations for (a) alleged fathers (b) alleged fathers but without DNA typing (c) incest cases and (d) missing persons. These programs have been named EasyPA, Easy Pant, EasyIN, and EasyMISS respectively [10].

It is a user friendly software for paternity testing (civil and criminal cases) and kinship testing (cases such as car accidents). Its user friendly capabilities allow the user to save results for future use and to make reports. This software is capable of handling any number of loci of the DNA profile of a person. Easy DNA is capable of carrying out a paternity test even when the alleged father cannot be typed. This means that the DNA of the father is not known.

Advantages

1. It is user friendly because it uses a built-in-pull-down manual to input genotypes.
2. A visual description of the relationships of the people involved in the test is provided by a pedigree diagram.
3. The output can be saved by the client for easy checking and for legal purposes.
4. It has a wide applicability because it is made up of four software applications. It can be used in paternity test where the DNA of the alleged father cannot be gotten, incest cases, inheritance dispute cases, missing cases and so on.
5. Apart from calculating the likelihood ratio, the posterior probability of paternity is also calculated.

Interface

1. Built-in-pull-down manual: this allows the genotypes to be easily inputted by the user

Into the software, hence the reason for its user friendliness.

2. Calculate: A tab responsible for calculating the likelihood ratio of the genotypes which the user enters in the case of a paternity or kinship testing.
3. Generate Output: the results are displayed and saved in a file on the computer system.
4. Pedigree diagram: This function allows the user to visualize the relationships of the people involved in the test

Interface for EasyMiss software is given in the figure below [10].

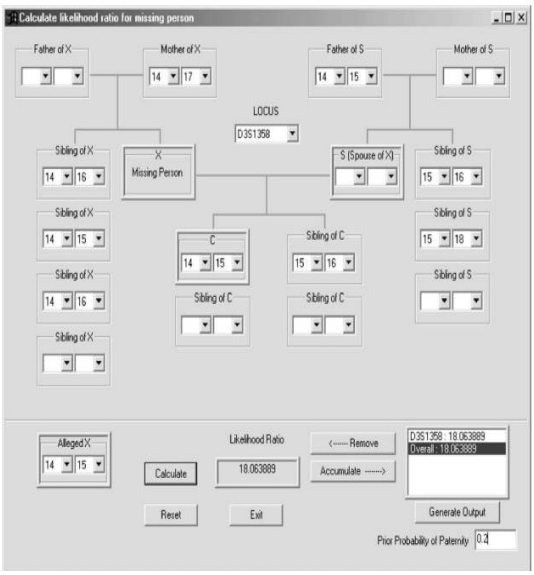


Fig. 3. The interface for EasyMISS software. It shows the two alleles of each person to be tested and the likelihood ratio is calculated using a prior probability of 0.2

IV. DISCUSSION

This literature gives a survey and analysis of three paternity test software. It has described the interface, advantages, basic functionalities and platforms of PedExpert, Genolab and EasyDNA. The Paternity test software have important features that distinct them from one another. These features range from the environment on which they are configured and operated, the purpose for their algorithms, their basic functions and mode of operation, interface, security mechanism and nature of the paternity case.

PedExpert operates on the Microsoft Windows environment. Ped Expert operates in the SMILE (Structural Modelling, Inference and Learning Engine) environment. SMILE.NET was created for the Microsoft.NET platform utilizing Microsoft Visual C++, but it can be used with any programming language supported by Microsoft.NET [6]. The database of PedExpert uses the paradox software which is commercialized by Corel Corporation. What distincts PedExpert from Genolab is the incorporation of Genolab with the mobile platform. Genolab uses a cloud-based service and so users can connect through their accounts to a shared network that provides high computing power and storage capacity. Genolab users can connect to the software through their mobile devices. Cloud computing is a technology which allows easy access to shared computing resources (e.g., networks, servers, storage, applications and services) that can be quickly retrieved [11]. EasyDNA is packaged software that comprises of four different software products. They are EasyPA, EasyPAnt, EasyIN and EasyMISS. EasyPA

is used to carry out a paternity test. EasyPAnt is used for complex paternity test in which the alleged father is missing and the DNA profile of the alleged father is not available. EasyIN is used for incest cases and EasyMISS is used for missing persons.

The algorithms used in PedExpert constructs a Bayesian network from the family pedigree. Originally algebraic calculations were used in complex situations where the DNA profile of the father cannot be analyzed because he is deceased or missing. This approach can be rather time consuming and error prone. Expert probabilistic systems known as Bayesian networks can be used to solve the complexity of using algebraic calculations. Its algorithms convert pedigrees and DNA information such as genotypes into Bayesian networks by using the SMILE.NET platform. The algorithm uses the relationship among the genetic information of the family members to create a Bayesian network. The algorithm also creates a table of conditional probabilities of each node in the Bayesian network and it fills the table with entries. This helps to reduce the time of analysis of complex cases in paternity tests. However, there was the need to solve complex paternity cases in sufficient time and with standard computer hardware. This led to the development of Genolab.

The algorithms used in Genolab are developed for cloud technology. Its algorithms are optimized to deliver the necessary computing power for complex paternity cases. In creating the pedigree design, it uses combinatorial deficiency analysis to determine all possible pedigree constellations. The kinship algorithm calculates the pedigree likelihood

The algorithms used by the software in the EasyDNA package are for user friendly capability of the software. It ensures the pedigree is designed using a pull down manual for easy visualization of the relationship of the people involved in the test. It also incorporates the Bayesian theorem to calculate the likelihood ratio. The Bayesian theorem involves using a prior probability to calculate the combined paternity index and likelihood ratio.

One important feature of Ped Expert is that it provides a platform for the identification of genetic mutation information. Genetic mutation can cause inconsistencies in the calculation of paternity index. After the creation of the tables of conditional probability, it considers the fact that there could be a possibility of gene mutation and so it provides the necessary information (mutation rate) and new nodes do not have to be added to the network. This allows the same Bayesian network to be used in different modes. This makes it a new tool for calculating the probability of paternity in paternity testing.

The basic functions of Genolab are paternity testing, kinship analysis and DNA evidence interpretation using a practical cloud based service. It is a user friendly software for paternity testing (civil and criminal cases) and kinship testing (cases such as car accidents). Also hundreds of kinship analysis can be carried out at once using Genolab's batch analyzer in cases of missing persons, mass graves or accidents. Its data security mechanisms are encryption of data and SSL- encrypted connections.

An important feature of EasyDNA is that it is a software for paternity and kinship testing which is user friendly and easy to use and understand. Its pedigree diagram uses a pull down manual that allows easy understanding of the relationship of the people involved in the test. The pull down menu also makes operations easy because you only need to click the DNA type ad gene frequency to calculate the paternity index. Also, the user does not need to have thorough knowledge of paternity testing to be able to use the software [12]. Its output can be saved by the client for later use.

PedExpert works by inputting the sets of marker Loci which make up the DNA profile into the software, then the pedigree can be created by selecting the pedigree ID field and the set of Loci required. The software constructs a Bayesian network from the pedigree. There is no need to create different network structure for each set of Loci because the network structure depends on the pedigree. This

is what separates PedExpert from other software that are time consuming because they have to create different Bayesian networks for each set of Loci typed. Next, the DNA typing data is inputted into the software. The data is exported to a pedigree- drawing software called a haplopainter which constructs a pedigree to ensure that all family relationships are correctly specified. Bayes theorem is used to calculate the paternity index with an apriori probability of 0.5. The result is then exported to a spreadsheet file. This whole process takes less than 5 minutes with the use of 30 microsatellite loci for complex cases.

Genolab carries out kinship and parentage testing through the following steps. Firstly, the DNA profile data of the people required are uploaded and stored as a project list in the user account. Next, hypotheses are created which represent different pedigrees showing relationships among the people required in the paternity test with the help of a graphical pedigree designer. Genolab contains some already constructed pedigree templates that can be used to reduce the time spent on designing pedigrees for various hypotheses. Lastly, the likelihood ration is calculated. It produces a result that shows the probability of every hypotheses defined and every marker of the genetic profile. The result can be exported or printed.

Easy DNA solves complex paternity cases were the alleged father is dead or missing by typing the DNA profiles of the child and mother at various locus, the hypotheses is generated and the likelihood ratio is calculated. The paternity result showing the probability of paternity is saved in a file.

V. CONCLUSION

It is important to make use of efficient techniques in reconstituting DNA profiles or genetic information of a person so that paternity testing of complex cases can be carried out in less time. This is why PedExpert incorporates the Bayesian network, to save the time it will take to carry out paternity test with no DNA typing. More time and space can also be saved by incorporating Cloud computing technology in paternity testing. Genolab's incorporation of cloud computing has saved time due to its high computation speed and fast access to stored information while performing operations on large amount of data at the same time. The advancement in technology allows reconstitution of DNA profiles of multiple persons at the same time which is a great effort in saving time and space. It is also very important for a software user to be able to connect to the software mentally with high understanding. EasyDNA incorporates user-friendliness into its design by improving the visuality of is pedigree design. This way the user understands the design properly even in complex paternity cases and will be able to make better hypotheses.

Conditional probability, which is gotten from the concept of Bayesian theorem is used in the calculation of the probability of paternity in the software surveyed. This because it uses a prior probability which indicates that the alleged father could be the biological father or not, and so there is no knowledge of the true hypothesis until after the test. A software for calculating likelihood ratio in paternity and kinship testing should be efficient in terms of time and space for simple and complex cases and it should be easy to use and understand.

REFERENCES

[1] K. Henningsen (1956). On the application of blood group testing to cases of disputed paternity in Denmark. Acta med Leg Soc (Liese) vol. 9 , pp 95 – 104.

[2] AABB (American Association of Blood Banks) 2006. Annual report summary for testing in 2006. Available at [http://www.aabb.org/Documents/Accreditation/Parentage_Testing_Accreditation_Progra m/rtannret06.pdf]. Accessed March 5, 2009.

[3] K. Apostolopoulos, E. Labropoulou, B. Konstantinos, S. Rhageed and E. Ferekidis (2002). Blood group in otitis media with effusion. ORL J Otorhinolaryngol Relit spec.64(6) pp 433 - 435.

[4] RD. Klein, DJ. Dykas and AE. Bale (2005). Clinical testing for revoid basal cell carcinoma syndrome in a DNA diagnostic laboratory. 7th ed, vol. 9 Genet med : pp 611 – 619.

[5] T. Aysim and G. Gulumser General (2009) characteristics of paternity testing applicants. TurkJ med sci 39(6) pp 923 – 931.

[6] R. Gomes, S. Campos and S. Pena (2009). PedExpert ; a computer program for the application of Bayesian networks to human paternity testing. Genetics and Moecular Research 8(1), ppv273 – 283.

[7] Maureen Young (2012). Paternity – Legal concerns and isues. Anylabtestnow. Pg 6 – 8.

[8] M. Frank Gotz (2015). Genolab – Forensic biostatistics with cloud computing. Forensic
Science International : Genetics Supplement Series, vol. 5 pp438 – 440.

[9] J. Buckleton, C. Triggs and S. Walsh (2005). Evidence Interpretation , CRC Press, Bocat Raton Florida.

[10] K. Fung, C. Yang, and W. Guo (2004). Easy DNA : user-friendly paternity and kinship testing programs. International Congress Series, vol. 1261, pp 628 - 630

[11] P. Mell and T. Grance (2011). The NIST definition of cloud computing. NIST special publication ; pp 800-145

[12] M. Takamiya, Y. Aoki and K. Dewa (2014). Programming Scripts for Simple and Complex Paternity Testing based on Open-Source Programming Language from the R Project. Journal of forensic science and criminology. Vol 2, Issue 2.

A survey of the local optimal sequence alignment characteristics of the Smith- Waterman Algorithm

Sharon Ikechi
Computer Science Department
University of Ilorin
Ilorin, Nigeria
Sharon.kechi@gmail.com

Bamidele Oluwade
Computer Science Department
University of Ilorin
Ilorin, Nigeria
deleoluwade@yahoo.com

Abstract— **Sequence alignment is a very important process in the field of Bioinformatics which can be carried out through local, global or semi-global alignment. Global alignment involves scanning the entire sequences for its comparison, however local alignment identifies similar regions from sequences of dissimilar length and this makes it preferable. The problem of sequence alignment has been worked on by applying computational algorithms to make local alignment easier and faster. Dynamic programming was incorporated to form the Smith-waterman algorithm. Furthermore, algorithms that make use of the heuristic approach have also been incorporated in local alignment to form local alignment search tools like BLAST and FASTA.The Smith-waterman algorithm is an efficient and effective algorithm for constructing the local alignment of sequences to generate optimal, accurate and precise results. It searches for subsequences or regions in the sequences that align and because these regions contain nucleotides that are located close to each other they have high structural and functional importance. This paper gives a survey of the local optimal sequence alignment characteristics of the smith-waterman algorithm.**

Keywords— Local optimal alignment; Global alignment; smith-waterman algorithm; scoring matrix; back tracing.

I. INTRODUCTION

Comparing two sequences involves determining the optimal alignment between them. The alignment can be global, if it contains all the characters of the sequence, it can be local if it contains substrings of the sequence and it can be semi-global if it contains prefix or suffixes of the sequence [1]. The purpose of Sequence alignment is to understand the structure and function of a newly discovered gene by comparing it with a gene whose structure and function is known. By comparing the genes through sequence alignment it will be possible for biological scientists to determine the similarities that appear in the genes. Global and Local alignment are the two types of sequence alignment process. Global alignment involves aligning all the nucleotides in both sequences as much as possible from the first nucleotide to the last nucleotide of both sequences. Local alignment involves aligning sub sequences within both sequences that align correctly giving matches [2]. Local alignment is the global alignment of substrings within the sequences [3].

The role of computer science in bioinformatics is to design algorithms that solve biological problems. An algorithm is a sequence of instructions that must be executed for a well formulated problem to be solved [4]. A well formulated problem is an unambiguous and precise statement that specifies the input, output and the method of translating the input statement to the output statement. In computer science, pseudocodes are mostly used to represent algorithms because they are more precise, less ambiguous and they hide irrelevant details. An algorithm is said to be correct if for every input instance of the algorithm, the correct output is generated. If this condition does not hold for every input instance, the algorithm is said to be incorrect. The number of steps the algorithm has to implement to generate the output is called its running time which depends on the size of the algorithm input. To determine the efficiency of an algorithm or its complexity, the running time of the algorithm has to be known.

In computer science, the big O notation describes the running time of an algorithm. Big O notation means the order of growth of the running time function, which is the leading term in the function. This means that a function of n is the limit or upper bound of the running time of an input of size n. An algorithm's running time is said to be quadratic if its big O notation is a quadratic function of n and the running time function will not grow faster than the quadratic function of n. the lower bound can also be determined showing that the running time function will not grow slower than the quadratic function of n which is denoted as omega (Ω). The upper bound of an algorithm specifies the algorithm in its worst case scenario, worst case means the

worst input instance of that algorithm, the worst behavior of that algorithm or the algorithm in its worst form. Similarly, the lower bound of an algorithm specifies the algorithm in its best case scenario, best case means the best input instance of that algorithm, the best behavior of that algorithm or the algorithm in its best form. In measuring the complexity of an algorithm, the worst case

scenario is used. It is just like analyzing the day ahead of you by imagining the worst things happening so that when they eventually happen you won't be disappointed.

There are various algorithm design methodologies such as brute force, branch-and-bound, divide and conquer, dynamic programming, machine learning, randomized algorithm and so on. However, this paper focuses on dynamic programming because the Smith-Waterman algorithm uses the dynamic programming technique. The popular divide and conquer approach which involves breaking a complex problem into smaller sub-problems and solving each sub-problem repeatedly could lead to waste of time when there are large number of sub-problems that must all be solved. Dynamic programming is a technique that avoids solving sub-problems that have already been solved or the solutions have already been computed to save a lot of time [4].

This research paper focuses on analyzing the result of computing the Smith-Waterman algorithm and how an optimal alignment can be generated from the algorithm.

II. SMITH-WATERMAN ALGORITHM

To obtain the structural, functional and evolutionary relationships between or among sequences, it is important to identify regions of similarities occurring in the sequences, this is done through a process known as sequence alignment. Sequence alignment is carried out by writing the sequences on successive lines across the page, with matching or same letters placed in the same column and unmatched or different letters placed in the same column as a mismatch or next to a gap as an insertion or deletion in the other sequence [2].

The Smith-Waterman algorithm was first proposed in 1981 by Temple F. Smith and Michael S. Waterman [5]. The algorithm is a modification of the Needleman-Wunsch algorithm. The Needleman-Wunsch algorithm is constructed to perform Global alignment on sequences and was developed by Saul B. Needleman and Christian D. Wunsch in 1970 (Needleman, 1970). The Smith-Waterman algorithm uses the dynamic programming technique to carry out sequence alignment that leads to an optimal alignment which is the best alignment possible. Smith and Waterman found out that regions in DNA structure that align well are of more biological significance compared to other regions that do not align well. They also found out that insertions and deletions in alignment are evolutionary changes and so they modified the Needleman-Wunsch algorithm to allow the identification of regions that align well and to accommodate insertions and deletions in the alignment. They were able to modify Global alignment using dynamic programming which led to the Smith-Waterman algorithm.

Smith-Waterman algorithm uses a substitution matrix and gap scoring system to find the best alignment of all possible alignments by giving scores to matches, mismatches and gaps. Local alignment searches for the extent of local similarity between sequences. The optimal alignment will have the best positive overall score out of the overall scores of all the possible alignments. In comparing two sequences through local alignment the following conditions must hold

1. The symbols in the first sequence should be aligned with the symbols in the second sequence.
2. All the symbols in the sequences should take part in the alignment process. When there are two sequences of dissimilar length, no symbols should be left out, the alignment should cover every symbol.
3. A symbol in one sequence can be aligned with a symbol in another sequence or a gap. A gap is a dash symbol which is used when there is no symbol to be used in that space because of the alignment arrangement.
4. A gap in one sequence cannot be aligned with a gap in another sequence.
5. A negative score must be given to mismatches.
6. The minimum value for in the matrix must be a 0. All negative numbers are replaced with a 0.

The aim of sequence alignment is to look for the most optimal alignment which identifies regions of high structural and functional similarities. Consider the example below showing the Local alignment process of two sequences, sequence 1 and sequence 2.

Sequence 1 : T C T C G A T
Sequence 2 : G T C T A C
Two ways of locally aligning sequence 1 and 2 are:
Alignment 1 : T C T C G A T
G T C T A C
Alignment 2 : T C T C G A T
G T C T A C

Match : it is indicated by the nucleotides in blue. It occurs when the same nucleotide occur in the same row in both sequences.

Mismatch: it is indicated by the nucleotides in black. It occurs when different nucleotides occur in the same row in both sequences.

Indel : it is a column that contains a space. There is no nucleotide in this column just a dash symbol or a gap.

Insertion : it occurs when a nucleotide in the first sequence is aligned with a space in the second symbol.

Deletion : it occurs when the space in the first sequence is aligned with a nucleotide in the second sequence.

How do we determine the optimal alignment of the two possible local alignments given in the example above? This is where the Smith-Waterman algorithm comes into play. The algorithm can be used to determine the best alignment out of all the possible alignment because the best alignment will have the highest match. Steps to perform a Local alignment are:

1. Obtain the sequences to be aligned and label them as sequence 1 and sequence 2.

2. Design a scoring matrix, it is called a scoring matrix because you put the scores for each nucleotide alignment in each box of the matrix. As you input the scores in the scoring matrix, mark the directions using arrows.

3. Compute the scores of the matrix using the following conditions:

i. Let s represent the first sequence and i represent the elements of sequence s.

ii. Let t represent the second sequence and j represent the elements of sequence t. iii. Let M represent the Matrix

iv. Represent sequence s as the elements in the column and sequence t as the elements in the row

v. Let the starting box which is $M_{1,1}$ be 0 and the score of the gap will be inserted for the first row and first column. This is done because gaps can vary for each alignment.

1.e M[i][0] = 0 and M[0][j] =0

4. To find M[i][j], which is the score for each column in the matrix, the scoring value for a box can come from 3 directions which are:

i. The box beside it (left)

ii. The box at its top

iii. The diagonal box next to it

The figure below shows the formula for calculating the scores for the matrix. Where S[si][tj] represents the score of a match or mismatch and g represents the score for a gap [6].

$$M[i][j] = \max \begin{cases} 0 \\ M[i-1][j-1] + S[si][tj] \\ M[i-1][j] - g \\ M[i][j-1] - g \end{cases}$$

fig.1. Formula for calculating the scores of the matrix

This means that the score of each box will be the maximum of three values which are gotten from the score in the box diagonal to it plus the value given for a match or a mismatch. It the letters in that box are the same then it is a match, otherwise a mismatch.

ii. The score of the box beside it (left) plus the value given for a gap. iii. The score of the box at its top plus the value given for a gap.

iv. If the three values gotten above are all negative, then the score for that box is reset to 0. That is why 0 is the first value in the matrix conditions. [7]

This is what differs the local alignment from the global alignment. The reason why the maximum score is gotten from these three values is because local alignment aims at looking for the best way it can be aligned to give the highest score. The application of dynamic programming in local alignment is in solving sub-problems like determining he values of the three conditions in the matrix and then calculating the maximum value.

Do not forget to draw the direction from which the calculated score is gotten from using an arrow.

1. Trace backing is carried out by starting from the highest number in the matrix and tracing it back to the starting node 0 using the arrows drawn. The highest value is the endpoint of the local optimal path.

2. After trace backing write out the sequences using the direction of the trace back arrows to know where to put a gap and a letter.

An Example of Local Alignment is given below.

To perform local alignment on the two sequences below using the Smith-Waterman algorithm, follow the steps given above.

Sequence 1 : T C T C G A T

Sequence 2 : G T C T A C

Using match = 2, mismatch = -1 and gap = -2

The flow of data to compute each alignment is given in the figure 1.

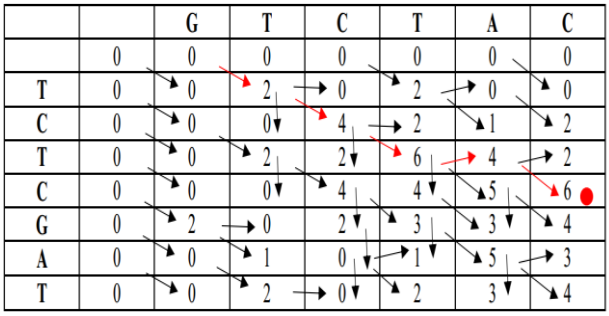


Fig. 2. A Scoring Matrix Design

the scores for the boxes in the first row and column are all 0.

The highest score is 6, back trace 6 to the starting point by following the arrow direction. It is shown by the red arrows. Shows that a character will follow and or shows that a gap will follow in the alignment. For example, in row 4 column 6, the shows that a gap will follow in the alignment.

Row 2, Column 2

i. Diagonal box = 0 + -1(T and G gives a mismatch) = -1
ii. Box beside it(left) = 0 – 2 (value for a gap) = -2
iii.Box at its top = 0 – 2 (value for a gap) = -2

The maximum number is -1 , but 0 is used because 0 is greater than -1.

Row 5, Column 7

i.Diagonal box = 4 + 2(C and c gives a match) = 6

ii. Box beside it(left) = 5 – 2 (value for a gap) = 3

iii. Box at its top = 2 – 2 (value for a gap) = 0

The maximum number is 6. (the process is repeated to get the scores for each box in the matrix)

Therefore the optimal local alignment for sequence 1 and 2 is

T C T – C G A T

G T C T A C

The incorporation of dynamic programming in the Smith-Waterman algorithm allows it to be broken down into sub problems. Solving the sub-problems leads to solving the bigger problem which is finding the local optimal alignment of sequences. Below are the pseudocodes for solving the two sub-problems.

Input : two sequences A and B.

Process: to calculate the optimal local alignment and score s of two sequences A and B, solve

1. Initialization of matrix
2. Smith-waterman alignment

Output : optimal local alignment and a score s.

Pseudocode for Smith-Waterman Algorithm is given below [8].

```
Pseudocode for Initialization of matrix
For i=0 to length (A)
F(i,0) ← d * i
For j=0 to length (B)
F(0,j) ← d * j
For j=1 to length (A)
For j=1 to length (B)
{
        Diag ← F(i – 1 j – 1) + S(Ai, Bj)
        Up ← F(i – 1, j) + d
        Left ← F(I, j – 1) + d
        F(i,j) ← max(Match, Insert, Delete)

}
Pseudocode for SW Alignment
For (int i=1; i<=n; i++)
For (int j=1, j<=n; j++)
Int =score[seq1.CharAt(i-1)] [seq2. CharAt(j-1)]
Int val= max(0, F[i-1][j-1]+s, F[i-1][j],d,F[i][j-1]
F[i][j] = val;
If (val == 0)
        B[i][j] = null;
  Elseif (val == F[i-1][j-1] + s)
        B[i][j] = new Traceback2 (i-1, j-1);
  Elseif (val == F[i-1][j] - d)
        S[i][j] = new Traceback2 (i-1, j);
  Elseif (val == F[i][j-1] - d)
        B[i][j] = new Traceback2 (i, j-1);
```

III. DISCUSSION

The example given above shows that it is possible to generate all the possible local alignments from the scoring matrix. However, the optimal alignment which is the alignment that has the highest score will always start with the largest number in the matrix. From the scoring matrix above, two local alignments among others can be generated which are:

Alignment 1 : T C T – C G A T

G T C T A C

this alignment has a score of 2+4+6+6=18

Alignment 2 : T C T C G A T

G T C T A C

this alignment has a score of 2+2+6=12

This shows that Alignment 1 has a higher score and is therefore the most optimal alignment. The Needleman-Wunsch algorithm used for global alignment and the Smith-Waterman algorithm used for local alignment differ in the results they produce after the alignment process. In pairwise alignment which involved aligning two sequences, the Needle-man Wunsch algorithm produces an optimal global alignment which contains gaps while the Smith-waterman algorithm will produce locally aligned subsequences. However, the two algorithms incorporate the use of dynamic programming to optimize the scores based on substitution matrices [9]. The local alignment process has been incorporated into various search tools such as BLAST and

FASTA due to its optimal performance and its accuracy and precision in results. BLAST means Basic Local Alignment Search Tool. It is a database searching tool that compares a query sequence with a database of sequences. It works by looking for short matches of about three letter words between the sequences. This process is known as seeding. After the seeding process, local alignment is carried out on the short matches as they extended in either direction, the score of the alignment is calculated using a scoring matrix. It is a sensitive algorithm that searches for more significant patterns in sequences [10]. Another search tool that incorporates local alignment in its operations is FASTA. FASTA means Fast Alignment. It works by looking for regions shared by the query and database sequences that have the highest degree or density of k-tuple sub-words. It uses a scoring matrix to identify the best or optimal regions and check for the possibility of joining regions with the use of gaps it then constructs an optimal alignment based on the Smith-Waterman algorithm [10].

In terms of accuracy and speed, the smith-waterman algorithm differs from BLAST and any other implementation of the smith-waterman algorithm. The results gotten from the use of the smith-waterman algorithm are more accurate than BLAST because it searches for matches that BLAST cannot and it covers the entire sequence. Searching for the optimal alignment using the scoring matrix allows the algorithm to cover every letter of the sequence, thereby leading to accurate results. However, it is more time consuming in its operation compared to BLAST.

IV. CONCLUSION

Determining the similarities that appear in genes is very important to biological scientists. The smith-waterman algorithm was developed to construct the local alignment of sequences to obtain the structural, functional and evolutionary relationships between or among sequences. The smith- waterman algorithm uses the dynamic programming approach in solving the problem of sequence alignment. Its pseudocode can be broken down into two sub problems. First is the initialization of a scoring matrix which involves comparing the sequences and filling the matrix with values and then the performance of back tracing to identify similar subsequences based on optimal scores. It uses a substitution matrix and gap scoring system to find the best alignment of all possible alignments by giving scores to matches, mismatches and gaps. The Smith-Waterman algorithm can be used as an optimal approach in the local alignment of sequences.

References

[17] P. Ananth and A. Ganesh (2014). Parallelization of Biological Gene Sequencing Technique with Optimized Smith Waterman Algorithm. International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 6, pp4552-4571

[18] Mount DM. (2004). Bioinformatics: Sequence and Genome Analysis (2nd ed.). ColdSpring Harbor Laboratory Press: Cold Spring Harbor, NY

[19] G. Clemen and R. Knut (2013). Pairwise Alignment. Cambridge university press pp. 19-31

[20] J. Neil and P. Paul (2004). An introduction to bioinformatics algorithms. Massachusetts

[21] F. Smith, Temple and S. Waterman, Michael (1981). "Identification of Common Molecular Subsequences" (PDF). Journal of Molecular Biology. 147: 195–197.

[22] A. Eduard., N. Juan and G. Dani (2007). Smith-Waterman algorithm. docencia.ac.upc.edu/master/AMPP/slides/ampp_sw_presentation.pdf <last retrieved on Oct. 13, 2016>.

[23] Shomu's biology (2013). Bioinformatics part 10 How to perform local alignment. https://www.youtube.com/watch?v=IatoWOsJ35Q <last retrieved on Oct. 10, 2016>.

[24] S. Biradar, V. Desai., B. Madagouda and M. Patil (2013). Comparative analysis of dynamic programming algorithms to find similarity in gene sequences. IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 08, pp 312-316.

[25] C. Abihlash and K. Rohitaksha (2014). A comparative study on global and local alignment algorithm methods. International Journalism of Engineering Technology and Advanced Engineering, Volume 4 Issue 1, pp 34-43.

[26] E. Donkor, N. Dayle and T. Adiku, (2014). Bioinformatics with Basic Local Alignment search tool (BLAST) and fast alignment (FASTA). Journal of Bioinformatics and Sequence analysis. vol6 (1)

.

Development of A Database for Fishes Using Semantic Web Technologies (Case Study of Lagos State, Nigeria)

Oyeniran Adedamola Olumide
Department of Computer Science
Lagos State University
Ojo, Nigeria
olumideoyeniran@yahoo.com

Adebayo F. Adekoya
Department of Computer Science
Lagos State University
Ojo, Nigeria
lanlenge@gmail.com

Ademola P. Abidoye
Department of Computer Science
Lagos State University
Ojo, Nigeria
ademola.abidoye@gmail.com

Bamidele A. Oluwade
Department of Computer Science
University of Ilorin
Ilorin, Nigeria
deleoluwade@yahoo.com

***Abstract*— This paper aims to collate data on fishes using Lagos state as a case study. These collated data are then stored using a database model associated with the Semantic Web – triplestore. The data gathered on these fishes are processed: adding their locations in Lagos state, their binomial nomenclature, common English names and also their images. In implementing the Semantic Web, an ontology is created that holds all the data on the fishes using the Web Ontology Language, OWL provided by the Protégé platform. This ontology can then be queried using an endpoint. This project makes use of an offline endpoint. The endpoint is a SPARQL Protocol and RDF Query Language, SPARQL. The SPARQL endpoint for this project is provided by Apache Jena Fuseki server. The ontology is then loaded into the endpoint for storage and querying. Several queries are run to determine the effectiveness and output of the entire system.**

Keywords*— *taxonomy; ontology; OWL; SPARQL; serializations

I. INTRODUCTION

Lagos state is a state blessed with enough water bodies. Among the advantages of having vast water bodies is also a vast variety of sea creatures of which the fishes will be considered in this paper. Due to the geographic location of Lagos state surrounded by enough water bodies, it can be said that Lagos state is a wetland with major water bodies in four of the five divisions of Lagos state namely: Badagry, Epe, Ikorodu and Lagos Island.

Collecting data on fishes and storing them for both present usage and future purposes is important for development and acquisition of knowledge. This paper will make use of the idea of the Semantic Web for proper processing, storage and presentation of the data collected on the fishes. The term 'Semantic Web' was coined by head of the World Wide Web Consortium and founder of the World Wide Web, Tim Berners-Lee to mean a web of data that can be processed by machines. According to [5], "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise and community boundaries".

A. Problem Statement

There are several issues that warrant the proposal of this paper topic whereby the paper will address:

i. Improper classification methods employed in the organization of gathered fish data. This paper would require a good knowledge of fishes and the steps involved in categorizing or classifying them, this is formally known as *fish taxonomy*.

ii. Presently, there is no computer-based fish classification system for fishes in Lagos waters.

The above-mentioned problems are the main issues this paper will aim at addressing. Maintaining the Integrity of the Specifications

B. Objectives

i. Study how the Semantic Web can be used to create a fish database.
ii. Design a computerized fish classification system.
iii. Develop and implement the designed system.
iv. Evaluate the performance of the developed system.

C. Literature Review

Reference [2] conducted research on ornamental fishes and fishing methods in a fishing village in Epe local government of Lagos State. The fishing tools used by the fishermen of that village were identified and also the challenges faced using those tools. The fishes caught also by the fishermen were identified.

D. Fish Classification

One of the problems of fish identification is improper classification. This leads to the term, taxonomy. Taxonomy is defined as the science of the description and classification of organisms, essential to the inventory of life on earth [6].

Taxonomy makes use of several ranks or levels and sub-levels to identify and classify organisms. This paper makes use of just the 7 ranks of taxonomy which are:

i. Kingdom
ii. Phylum
iii. Class
iv. Order
v. Family
vi. Genus
vii. Species

These ranks help in the proper identification of fishes.

II. SEMANTIC WEB TECHNOLOGIES

The term "Semantic Web" was first coined by Tim Berners-Lee to mean a web of data that can be processed by machines. The Semantic Web also known as Web 3.0 is deemed by many to be the future of the internet. "The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation." It is a source to retrieve information from the web (using the web spiders from RDF files) and access the data through Semantic Web Agents or Semantic Web Services [4].

The Semantic Web is implemented using several technologies:

RDF – Resource Description Framework.

OWL – Web Ontology Language.

RDFS – Resource Description Framework Schema.

SPARQL – SPARQL Protocol and Resource Query Language.

RDF is a framework created for representing data in the form of triples.

These triples as the name signify consist of the parts: Subject, Object and predicate.

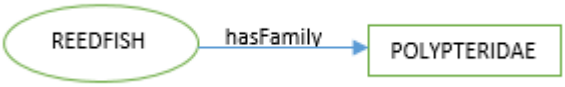


Figure 2.1: RDF triple

RDF triples in file format are known as serializations. Several serialization syntaxes are

i. RDF/XML
ii. TTL (TURTLE)
iii. N3

RDF/XML is the first standardized serialization syntax but TTL is gaining popularity because it is easier to use. RDFS and OWL are technologies or languages that are used to define ontologies, classes, subclasses, domain, range and several other semantic properties. An ontology is a formal naming and definition of the types, properties and their interrelationships of the entities that exist for a particular domain of discourse [1]. SPARQL is the query language used in querying the triples defined in serializations.

III. DEVELOPMENT OF DATABASE FOR FISHES

Developing the database for fishes requires three steps:

i. Gathering data on fishes in Lagos State. These data include: the name, its classification, location and image.
ii. Create an RDF serialization using the garnered data. The Web Ontology Language, OWL was used to create the ontology which consists on the fish data and other data that will help in linking data together.
iii. Query the ontology.

Creating the ontology would require sufficient knowledge in XML, but this paper makes use of Stanford's Protégé in creating the ontology. This software then creates a .owl serialization with the selected syntax in this case, XML.

The standard querying language for Semantic Web data is SPARQL Protocol and RDF Query Language, SPARQL. The model of database used with Semantic Web is the TRIPLESTORE. This triplestore provides a SPARQL endpoint (interface for inputting queries and obtaining results) both locally and remotely.

The fish ontology consists of two classes all subclasses of the class – Thing. The two classes are:

i. Fish
ii. Value:
    - Taxonomy
    - Others

The value class has two subclasses where the Taxonomy class contains the taxonomic description of the fishes. The 'others' class contains other fish properties such as

i. Habitat
ii. Location
iii. Image

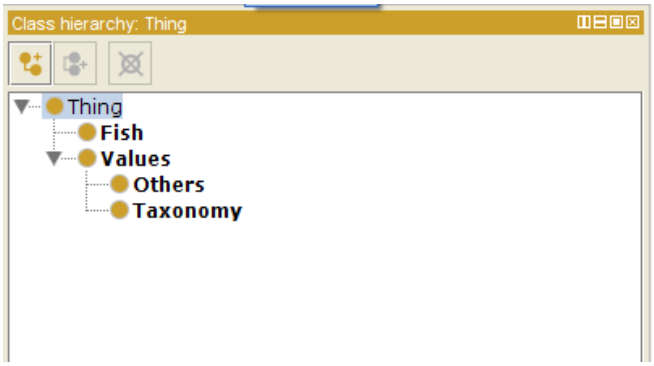


Figure 3.1: Ontology Class

The fish ontology consists of object properties and individuals and data properties. The object properties are properties that show the relationship between the subject and object of triples. In fig 3.1, the object property of the triple was 'hasFamily'. As with classes, the defined object properties are subsets of the object property 'topObjectProperty'.

The following object properties were defined to elucidate the relationships existing between the fishes and their specific data.

i. hasKingdom

ii. hasPhylum

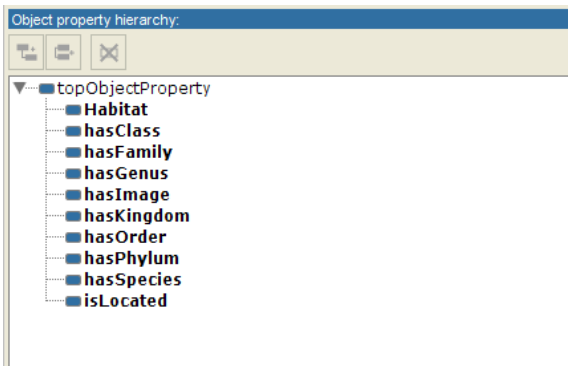


Figure 3.2: Object Properties

iii. hasClass

iv. hasOrder

v. hasFamily

vi. hasGenus

vii. hasSpecies

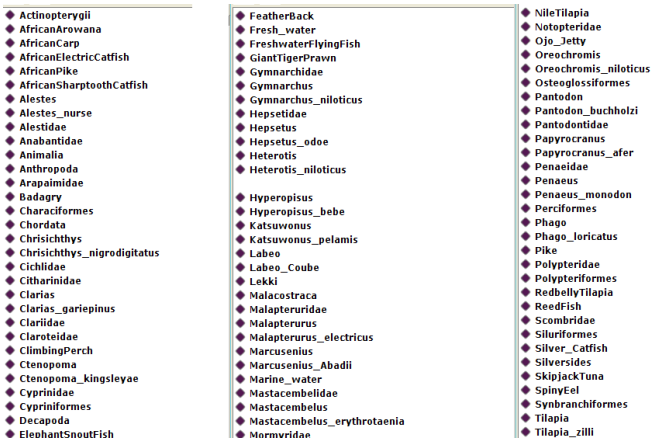


Figure 3.3: Ontology Individuals

viii. isLocated

ix. Habitat

x. Image

The individuals of the fish ontology are the entities that are assigned the above-mentioned object properties. The individuals are listed in the fig 3.4.

IV. DISCUSSION

Some queries are run on the developed system to determine if it is working as it should.

A. Identifying All Fishes That Belong to Kingdom 'animalia'

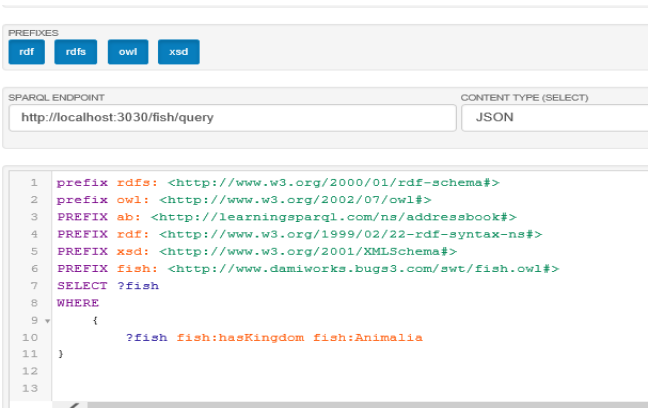


Figure 4.1: Kingdom Query

This query specified below is to determine the fishes that belong to the kingdom-animalia

SELECT ?fish

WHERE { ?fish fish:hasKingdom fish:Animalia }

The variable ?fish stores the fish English names that have the object property 'hasKingdom' and value Animalia. The

hasKingdom property can be specified by stating its full URI that is

http://www.damiworks.bugs3.com/swt/fish.owl#hasKingdom.

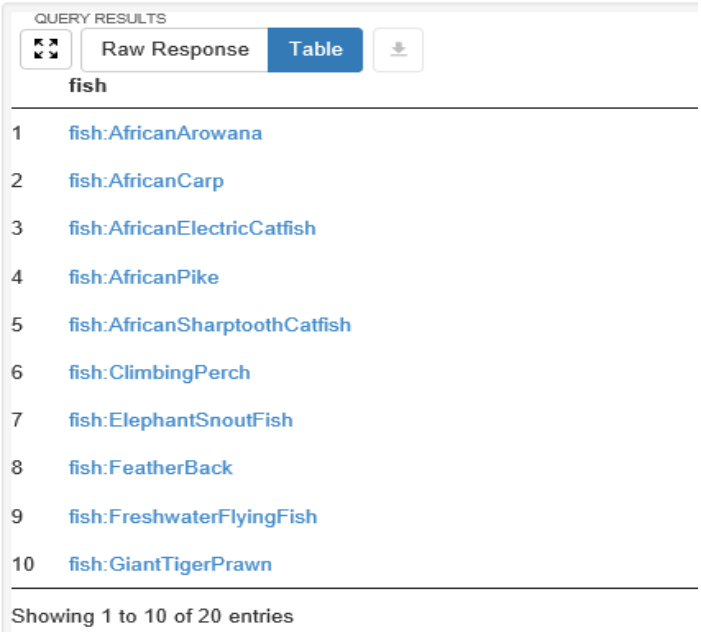


Figure 4.4: Epe Fish Query Result

This will take time and therefore the use of a prefix, in this case 'fish' that was assigned the URI http://www.damiworks.bugs3.com/swt/fish.owl# is appended to the property desired, in this case, that property is hasKingdom and also Animalia. The result for the above query is given in fig 4.2.

B. Identifying The Fishes That Are Located In The Epe Local Government

The following query provides the fishes in Lagos state located in Epe local government for this test:

SELECT ?fish

WHERE { ?fish fish:isLocated fish:Epe }

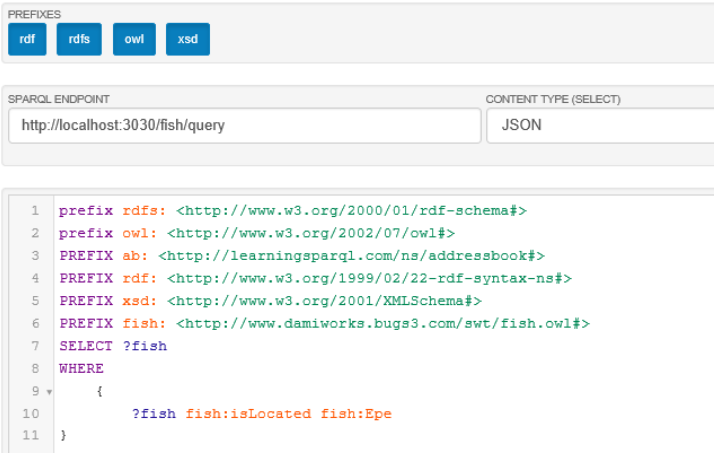
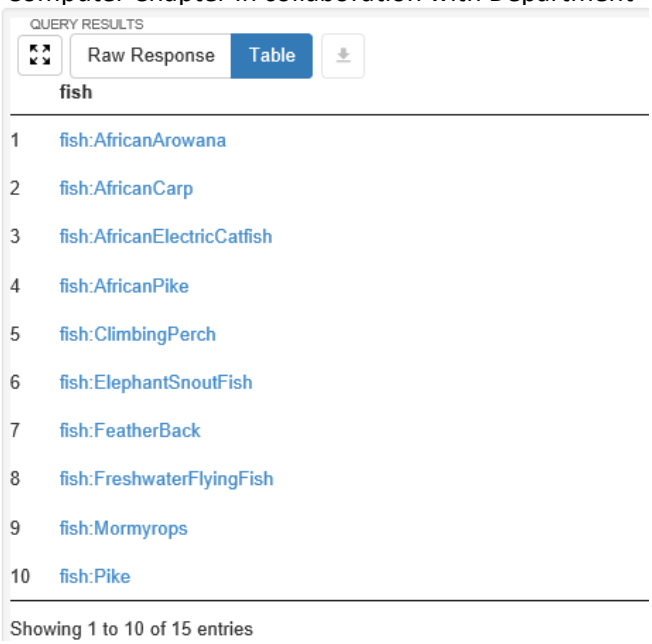


Figure 4.3: Epe Fish Query`



The result of the query above is displayed in fig 4.4.

CONCLUSION

The Semantic Web or Web 3.0 is the next generation of the internet. It is not so different from the World Wide Web (Web 2.0). Instead, it adds more functionality and meaning to the web. Hence, the name 'Semantic' Web. This can be achieved via several technologies such as RDF, SPARQL, XML, RDFS and OWL.

The Semantic Web technologies were used to create a database that stores data about fishes in Lagos waters. The fish data were gathered from students of fisheries department, Lagos State University and from the work of [2].

The Web Ontology Language was used in creating the fish ontology for the paper. The Jena Fuseki Server was used in the definition of the triplestore where the triples were loaded and stored and queried. The Fuseki server also provided an interface for querying this data. The fish ontology has been made available online and can be used in the description of other data(triples) just like the Friend of a Friend (FOAF) and the Dublin Core ontologies.

There are several other fishes which were not properly processed and therefore were not included among the 20 distinct species of fishes in Lagos waters that were inserted into the triplestore.

The images of the fishes were unable to be displayed in the ontology created. The images can only be viewed by visiting the link that comes as a result of the image query of each fish.

The fish database created has not yet been made public.

This paper can be improved by working on the limitations listed above. The Semantic database or triplestore can contain more and more triples of more fishes in Lagos waters making it wide enough for use in the Fisheries department of Lagos State University and other universities in Lagos state, other states of Nigeria and the world at large.

The SPARQL endpoint offered locally by the Jena Fuseki server can be made public just as the ontology has to allow querying of the fish data globally.

In conclusion, the Semantic Web is an advancement in technology that aims to make sharing of information a lot easier. The Semantic Web can be applied to any area of study.

It is the dream of Tim Berners-Lee that the Semantic Web be accepted and employed in the implementation and improvement of the World Wide Web [3]. This paper has gone a considerable distance in realizing that dream.

References

[1] Gruber, T (2008). Ontology. Encyclopedia of Database systems. Liu, Ling; Özsu, M. Tamer (Eds). ISBN 978-0-387-49616-0.

[2] Mbawuike B. C. & E. Ajado (2005). Assessment of Ornamental fish species and fishing methods in Ibiajegbende, Lagos state, Nigeria. Retrieved from the African Journal of Applied of Zoology & Environmental Biology, 2005. Vol 7, 23-27.

[3] Tim Berners-Lee, Nigel Shadbolt & Wendy Hall (2006). The Semantic Web Revisited. IEEE Intelligent Systems, pp 96-101.

[4] Tim Berners-Lee, Ora Lassila & James Hendler (2001). The Semantic Web. Scientific American – SCI AMER, vol 208, no 5, pp 34-43.

[5] W3C (2012). "XML and Semantic Web W3C Standards Timeline". 2012-02-04.

[6] Wagele Johann-Wolfgang (2005). Foundations of Phylogenetic Systematics. 2nd, revised edition..

Driving Electronic Health Record System Implementation in Nigeria: A Proposal

Olayiwola W. Bello[1] Nasir Faruk[2], Abdulkareem A. Oloyode[2], Segun I. Popoola[3]
[1]Department of Information and Communication Science, University of Ilorin, Ilorin, Nigeria
[2]Department of Telecommunication Science, University of Ilorin, Ilorin, Nigeria
[3] Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria

Email: laibello, faruk.n, oloyede.aa{@unilorin.edu.ng} segun.popoola@stu.cu.edu.ng

Abstract— **Information technology can be used to drive a sustainable and veritable health delivery system. This can be achieved through the introduction of various health information systems of which Electronic Health Record system is paramount. However, the implementation of electronic record system in the health sector in Nigeria is yet to have a wide uptake due to implementation challenges. This paper highlights some of these challenges which include poor coordination among stakeholders, ad hoc, piecemeal basis implementation by foreign software developers and applications designed to solve specific problems among others. The benefits of EHR and open source options available are discussed, and the documented implementation of EHR in the country presented. Also, an EHR implementation framework, based on a community of academics, technology enthusiast, open source developers and intended users of the application, with support from regulatory bodies is proposed. This can accelerate the attainment of the health component of the millennium development goals for Nigeria as well as the Vision 20:2020.**

Index Terms- Heath Information System, Electronic Health Record, Open Source, EHR Implementation, Nigeria

I. INTRODUCTION

The health of a nation can signify the wealth of the nation going by the popular saying that "health is wealth". Similarly, the power of information technology to translate processes and procedures into success is a well-known modern day "e" phenomenon. Aggregating these two positons presupposes that a synergy between information technology and health delivery system has the capability to bring improvement to a nation's general outlook. Collaboration between the health and ICT sectors is key to development. This is recognized by the both World Health Assembly Resolution 58.28 of 2005 and ITU World Telecommunication Development Conference Resolution 65 of 2010 [1]. Same position is corroborated by the opinion of Levingston [2] that:

> "Medicine is an information-rich enterprise. A greater and more seamless flow of information within a digital health care infrastructure, created by electronic health records (EHRs), encompasses and leverages digital progress and can transform the way care is delivered and compensated. With EHRs, information is available whenever and wherever it is needed"

World health organization (WHO) defines eHealth as the use of information and communication technologies (ICT) for health. In its broadest sense, eHealth is about improving the flow of information, through electronic means, to support the delivery of health services and the management of health systems. This can range from informational, educational, and commercial, to direct services offered by healthcare organizations, professionals, and consumers themselves [3]. Simply stated, e-health is making healthcare more efficient, while allowing patients and professionals to access and manage data in ways that were previously impossible. The growing impact of eHealth is reflected in the National eHealth Strategy Toolkit and it is bringing advancement to the delivery of health care around the world today. Similarly, it is making health systems more efficient and more responsive to people's needs and expectations [1]. It is also opined that technological advances, economic investment, and social and cultural changes are contributing to the realization that the health sector must now integrate technology into its way of doing business [1].

As an appendage to the general eHealth concept, the implementation of Electronic Health Record (EHR) system seem to still elude most developing countries including Nigeria. As Identified by the World Health Organization (WHO), part of the challenges facing the Nigerian health system is an inadequate health information system for monitoring and analysis of health indicators [2]. Indeed, the functionality of a well-designed EHR system can bridge this gap. This is despite that, over a decade ago, it was being taunted that we (the world) were near the tipping point where one can expect a steady rise in the number of health information system implemented and in their intensity of use in different settings, especially by healthcare providers at the point of care [5]. Apparently, this is yet to manifest in present day Nigeria as most health facilities are still living in the manual era of health records documentation.

Nigeria eHealth policy has been published since 2007, it includes a three-year strategic plan to reform Nigeria's National Health Management Information System. Emphasis is placed on improving data collection capabilities and

systems integration in order to support planning, monitoring and evaluation of health services. This includes the Public Health Care information system, which monitors health indicators such as infectious disease prevalence, maternal and child health outcomes, and rates of immunization [6]. However, Nigeria was not in the list of countries having eHealth profile for 2015. This indicates that close to nothing has been achieved in this direction.

Options available for acquiring EHR systems cuts across proprietary, development from the scratch and Open source /Free Software. Currently, there are many types of open source health information systems (HIS) in use. These systems can afford many health care providers the opportunity to use electronic health records without having to incur large costs. Also, they are able to make modifications to the software in order to suit their needs [7]. While debate is still ongoing on the proposal for a nationwide electronic health record system in Nigeria, this paper presents a community driven framework for the implementation of a nationwide EHR. Section 2 examines health information systems while x-raying the benefits and implementation challenges of EHR. Section 3 presents a survey review of some free and open source EHR applications. The concept of open source and free software development was discussed alongside a survey of open source options, African initiatives and a more detailed Nigerian experience. Section 4 presents the proposed framework before conclusion.

II. HEALTH INFORMATION SYSTEMS

Health Information Systems can be grouped into three distinct groups and these are: electronic medical records (EMR), electronic health records (EHR), and personal health records (PHC). Electronic health records (EHR) is a collection of health information about individual patients and populations in electronic format [8]. ISO considers EHR to be an overall term for "a repository of information regarding the health status of a subject of care, in computer process able form" [9]. It includes medical history, contact, hospitalizations and insurance information, family history, list of medications taken or currently prescribed, and allergies. These can all be shared with various hospitals or doctors' if necessary [7]. EHRs are designed to make it easier to share and use a patient's health care record and they have the propensity to reduce medical errors, improving quality of care, conserving physician time, sharing patient information among healthcare practitioners, and ensuring workflow efficiency [10].

An electronic medical record (EMR) is an electronic record of a patient's health information that can be created, gathered, and managed by doctors and their staff within one healthcare organization whereas an electronic health record is a patient's health information that adheres to national standards in the electronic format. It (EMR) can also be created and managed by doctors and their staff across more than one health care organization. A personal health record is an electronic record of an individual's health information that, in similarity with the electronic health record has to conform to national standards but is controlled by that individual [11, 12, 13].

EHR Benefits and Implementation Challenges

The benefits of EHRs have been examined by researchers [12, 14]. A triad perspective including the clinical, organizational, and societal outcomes were considered. From the clinical point of view, associated benefit includes; enhancements in the quality of care, minimizing medical errors, and other improvements in patient-level measures that describe the appropriateness of care. Organizational outcomes, on the other hand, have included such items as financial and operational performance, as well as satisfaction among patients and clinicians who use EHRs. Lastly, societal outcomes include being better able to conduct research and achieving improved population health.

Not only will the implantation of EHR speed up doctor patient attention, the following can also be achieved as documented by Menachemi, and Collum [12].

- Improve the accuracy and quality of health data record
- Enhances physicians' access to a patient's healthcare data
- Enables a seamless patient health record sharing for immediate and continuing care
- Improves care quality as a result of having health information immediately available at all times for patient care
- Improve the efficiency of the health record service
- Contain healthcare costs

All, these including the need to be part of the new world order in global healthcare delivery makes the adoption of EHR an attractive endeavor. The health system in Nigeria is both government and private driven. From the government side, a three tier system involving the Federal, State and Local Governments is noticeable. The private players are either private health care providers or Non-Governmental Organizations. Pantuvo et al., [15]. shows that the drivers for adoption of a nationwide EHR include the need to report data; improve patient safety, improve work place efficiency; comply with government reforms aimed at reducing the cost and increasing access to health services. It was however noticed that corruption, poor coordination among stakeholders, and lack of constant supply of electricity form some of the barriers to a successful implementation of a nationwide EHR. In the position of Moodley et al. [16], the limitation of health information systems infrastructure in most African countries has to do with poor implementation, caused by ad hoc, piecemeal basis implementation by foreign software developers and designed to solve specific problems. Little attention is said to be paid to how these components can fit into an integrated nationwide health information system and interoperate with other components.

Furthermore, WHO [17]. identified that the obstacles are not necessarily the technology available but more of technical support, in the form of technical expertise and the associated cost of migrating to an electronic system. This is not to foreclose the paucity of healthcare funding. Additionally, medical practitioners and health professionals' resistance to a

change from manual to electronic documentation is a cause for concern. Factors considered critical to a successful implementation of a nationwide EHR include enforceable legislation, a trained and motivated workforce, and significant and sustainable funding [15]. While it is one thing to have the HER in place, its utilization is another issue all together. As identified in the case of implementing Electronic Health Records in three East African countries, the system's use and sustainability varied depending on who controlled clinic funding [18]. It was also posited that successful EHR use and sustainability were enhanced by local control of funds, academic partnerships (mainly by leveraging research funds), and in-house technology support. From this experience, it might be necessary to factor in some of the aforementioned towards a sustainable system implementation.

III. OPEN SOURCE AND FREE SOFTWARE DEVELOPMENT

Going by the huge number of software products leveraging the OS concept, software and its development is attracting considerable attention from the open source software paradigm. OSS is unique in its production, modification, and distribution. The concepts of free and open software are closely related. A free software according to the Free Software Foundation[1] (FSF), is a software in which the user is free to share, study and modify (fsf.org). On the other hand, the Open Source Initiative[2] (OSI) is concerned with making available the source code and the distribution terms of open-source software are also clearly spelt out (opensource.org). This is to allow for a wider base developer interaction towards a more functional application. These concepts present adequate opportunities to developers to fast rack software development process with high end functionality. Bahamdain [19], asserted that the public availability of OSS will improve quality in terms of service level, productivity and end-user satisfaction. Similarly, it has been observed that OSS increases motivation and performance, more flexible in terms of productivity and it allows faster bug detection and error resolution compared to closed source software [20].

Notably, the new dawn being experienced in social computing as facilitated by global improvement in internet penetration has tremendously boosted the opportunities available for OSS projects. This gives a higher opportunity to access an even larger crowd for development [21]. Platforms through which this new form of collaborative development thrives include GitHub[3], BitBucket[4], Sourcetree[5], Sourceforge[6] and Google Code[7] among others. Going by the volume of projects being hosted on them, these platforms are indeed fast gaining grounds and with the list of numerous EHR/EMR available through open source offering, one can say OSS is gaining impact this area.

Open Source Electronic Health Record Systems

Significant efforts have been made in the OS community to drive the popularity of electronic health records. This can be substantiated by the array of development platforms available. Developers can harness the potentials of these existing application to develop a functional system with the required customization and additional functionality. Table 1 presents some selected open source applications.

These applications are made available at zero cost of acquisition and developers can also have aces to the entire source code the source codes. Virtually all the listed applications are still widely supported by a large community of developers, which make it an attractive option for developers. Interestingly, most of the applications are also designed to run on cross-platforms.

Table 2: *Electronic Health Record Implementation in some African Countries*

SOFTWARE/URL	License/Offering	Operating System
HOSxP/ hosxp.net/joomla25	GNUGPL/ free and open	Server: Linux Client: Windows
FreeMED / freemedsoftware.org	GPLv2 /free and open	Cross-platform
OpenEMR/ www.open-emr.org	GNUGPL / free and open	Cross-platform
GNUmed wiki.gnumed.de/	GNUGPL /free and open	Cross-platform
World Vista http://worldvista.org/	GPLv2 / free and open	Server: Cross-platform Client: Windows
Clear Health/ clear-health.com/	GNUGPL/ Not Stated	Cross-platform
OpenMRS / http://openmrs.org/	OpenMRS Public License /free and open	Cross-platform
OSCAR https://oscar-emr.com/	GPLv2 /free and open	Cross-platform
Zerps https://github.com/chrisekelley/zeprs	Apache Software License/ free and open	Windows
THIRRA https://sourceforge.net/projects/thirra/	MozillaPublicLicensev1.1 /free and open	Cross-platform
GNUHealth http://health.gnu.org/	GNUGPL /free and open	Cross-platform

Table 2 presents some African initiatives in the area of health information system using open source platforms. Some instances of the use of these applications are documented in some countries across Africa.

Table 2: *Electronic Health Record Implementation in some African Countries*

Country	Development Platform	Application
South Africa	Open Source (iDART)	HIV care
Ghana	Open Source (MGV-net)	Vital registration
Mozambique	Open Source (Open MRS)	HIV care
Rwanda	Open Source	HIV Care
Nigeria	Open Source (Open MRS)	Maternal & reproductive health,
Botswana	Open Source	HIV care
Cameroon	MEDCAD) (In-house Developed	Primary care
Zambia	ZEPRS (In-house Developed)	Public obstetric clinics, Teaching Hospital

However, most of the implementation were targeted towards specific application. As presented in Table 2, their use in HIV care is more predominant. While some countries in Europe have attained the level of having a shared countrywide EHR, African countries are still struggling to achieve this. It is important to note that Cameroon and Zambia has been able to develop separate medical records information system, developed in-house without the use of open source applications. These applications are also limited in their application as well as in coverage.

In Nigeria, available data suggest that OpenMRS is one of the popular open source option for health records system. It is also widely used across the world and well embraced in Africa. Despite the identified challenges to the implementation of EHR, available data from openmrs.org shows 11 health institutions running on the open source platform. These health institutions are government owned and cuts across the three tiers of Nigeria's heath delivery system. Six of the sites are projects of the same organization Institute of Human Virology, Nigeria but with no evidence of ability to share authorized health data online.

Table 3: Implementation of OpenMRS in Nigeria

Facility	Management	Last Updated
University of Abuja teaching Hospital (UATH)	Institute of Human Virology, Nigeria	6/7/2016
Benue State University Teaching Hospital (BSUTH)	Institute of Human Virology, Nigeria	6/7/2016
Federal Medical Centre Keffi	Institute of Human Virology, Nigeria	6/7/2016
PHC Lafia East, Nasarawa State	Not Available	19/4/2016
PHC Masaka	Not Available	19/4/2016
Mararaba Gurku Medical Centre	Institute of Human Virology, Nigeria	12/9/2015
Institute of Human Virology, Nigeria	Institute of Human Virology, Nigeria	28/10/2015
General Hospital Okwe	Institute of Human Virology, Nigeria	12/9/2015
General Hospital Minna - Niger State	Not Available	6/7/2016
Federal Medical Centre FMC Katsina	Not Available	19/1/2015
General Hospital Gaya, Kano	Not Available	19/1/2015

IV. THE PROPOSED IMPLEMENTATION FRAMEWORK

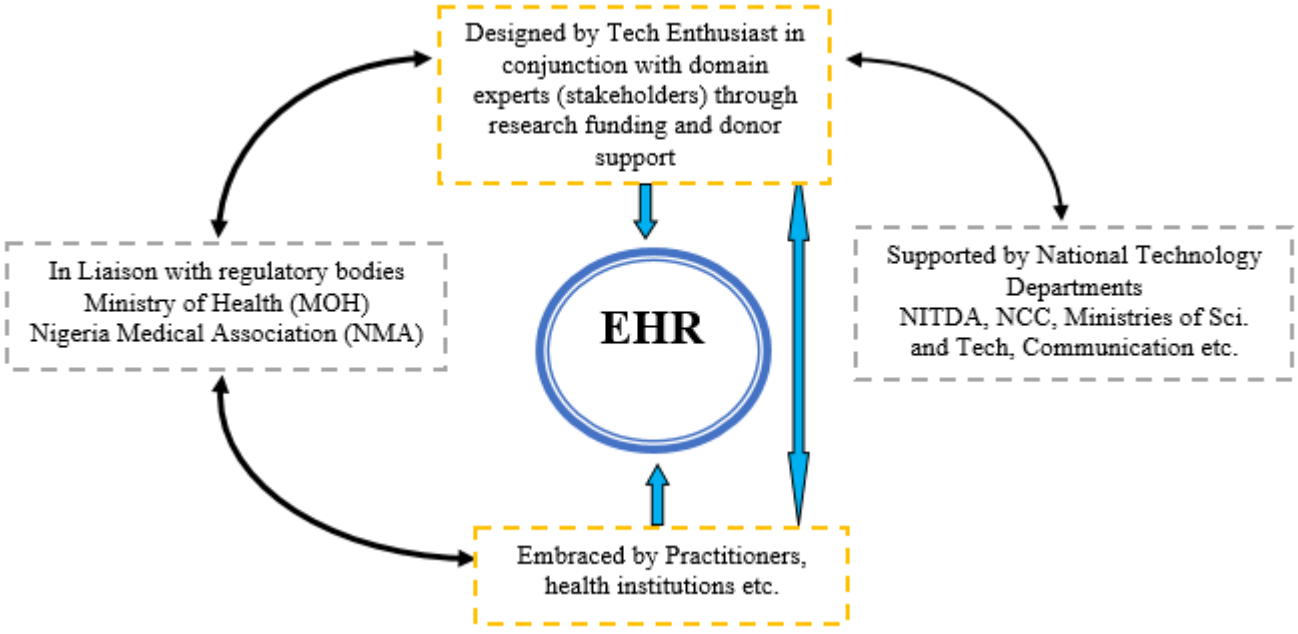


Figure 1: Community Driven HER Implementation Framework

The increasing popularity of not for profit organization coupled with the funding provided by donor organizations and access to research funding creates an enabling environment for a bottom-up approach to community driven initiatives in information system development. Interestingly, approaches to health promotion could be top-down or bottom-up. While the top down is a government driven initiative, the bottom-up is predicated on community based efforts. Evidence abound in the literature as to the paradigm shift towards 'bottom-up' approach in the realm of health promotion [22]. It is on this premise that we propose a community driven bottom-up approach to the implementation of EHR. This approach has at its core, the intended users of the system, immediate stakeholders and a community of developers and information technology enthusiast facilitated by academic researchers. Figure 1 presents the proposed framework. This framework drives EHR implementation through academics through research funding and a well-articulated collaboration with domain experts and other stakeholders. This effort will also incorporate regulatory bodies to accommodate their provisions. Support of national technology departments is also integrated. With this perspective to the implementation of EHR, a proper coordination among stakeholders can be achieved. This will ensure a proper articulation of requirement and standards and minimize piecemeal basis implementation, which will therefore erode the implementation by foreign software designed to solve specific purpose. A strong technical support could be achieved as the project is driven by home grown developers, facilitated by academic partnership. This implementation framework is proposed based on the synthesis of issues established in the literature, forming part of the major challenges to EHR implementation as discussed earlier.

V. CONCLUSION

Electronic health record system has the potential to turn around the health fortune of Nigeria. This will greatly impact on the country's healthcare service delivery capacity, disease surveillance and control. The availability of well-developed and adequately supported open source application can be adapted and redesigned to specificity. This approach will allow for rapid application development and allow for inputs from diverse developers with varying technical knowhow. More importantly, it could be the most cost effective way of implementation of EHR. Generally, it has the propensity to mitigate most challenges associated with past implementations. Findings from this paper show some instances of electronic record system in Africa in general and Nigeria in particular. However, integrating the isolated implementation into a nationwide shared information system has not been attained. With a bottom-up approach driven by an ecosystem of academics, technology enthusiast, open source developer and intended users of the application, a community driven EHR implementation framework s proposed. This can accelerate the attainment of the health component of the millennium development goals for Nigeria as well as the Vision 20:2020.

REFERENCES

[1] WHO & ITU (2012). National eHealth Strategy Toolkit. Retrieved 11/7/2016 from http://www.itu.int/dms_pub/itu-d/opb/str/D-STR-E_HEALTH.05-2012-PDF-E.pdf

[2] Levingston, S. A. (2012). Opportunities in physician electronic health records: A road map for vendors. *Bloomberg Government*. Accessed from http://www.healthit.gov/providers-professionals/why-adopt-ehrs.

[3] Alsulamea K., Khalifab M., & Househ M (2015). Health status in Saudi Arabia: A review of current literature Health Policy and Technology. (5) 204–210

[4] WHO (2009). Country cooperation strategy at a glance: Nigeria. Retrieved 15/7/2016 from http://www.afro.who.int/en/nigeria/country-cooperation-strategy.html

[5] Berner, E. S., Detmer, D. E., & Simborg, D. (2005). Will the wave finally break? A brief view of the adoption of electronic medical records in the United States. Journal of the American Medical Informatics Association,12(1), 3-7.

[6] WHO (2010). Global Observatory for eHealth. Retrieved August 25, 2016 from http://www.who.int/goe/policies/countries/nga/en/

[7] Jones, B., Yuan, X., Nuakoh, E., & Ibrahim, K. (2014). Survey of Open Source Health Information Systems. Health Informatics-An International Journal, 3, 23-31.

[8] Wan, Y. (2010). Application of HER in Health Care, Multimedia and Information Technology (MMIT), 2010 Second International Conference on, vol. 1, no., pp.60,63, 24-25.

[9] ISO (2004). Health informatics: Electronic health record - Definition, scope and context. Draft Tech Report 2004:03–16. ISO/DTR 20514. available at https://www.iso.org/obp/ui/#iso:std:iso:tr:20514:ed-1:v1:en

[10] Thakkar, M., & Davis, D. C. (2006). Risks, barriers, and benefits of EHR systems: a comparative study based on size of hospital. Perspect Health Inf Manag, 3(5), 1-19.

[11] Wager, K. A., Lee, F. W., & Glaser, J. P. (2009). Health care information systems: a practical approach for health care management. John Wiley & Sons.

[12] Menachemi, N and Collum T. H. (2011). Benefits and drawbacks of electronic health record systems. *Risk Management and Healthcare Policy*. (4), 47–55.

[13] HealthIT (2015). Benefits of Electronic Health Records (EHRs). Retrieved from 17/7/2016 https://www.healthit.gov/providers-professionals/benefits-electronic-health-records-ehrs

[14] King, J., Patel, V., Jamoom, E. W., & Furukawa, M. F. (2014). Clinical benefits of electronic health record use: national findings. *Health services research*, *49*(1pt2), 392-404.

[15] Pantuvo J. S., Naguib R. N. G., Wickramasinghe N. (2011). Towards Implementing a Nationwide Electronic Health Record System in Nigeria. International Journal of Healthcare Delivery Reform Initiatives, 3(1), 39-55, January-March 2011

[16] Moodley, D., Pillay A. W., Seebregts C. J. (2012). Researching and Developing Open Architectures for National Health Information Systems in Developing African Countries. In: Foundations of Health Informatics Engineering and Systems Ediors: Zhiming Liu, Alan Wassyng. Lecture Notes in Computer Science Volume 7151 2012 ISBN: 978-3-642-32354-6 (Print) 978-3-642-32355-3

[17] WHO (2006). Electronic Health Records: Manual for Developing Countries. ISBN 92 9061 2177. Accessed 10/10/ 2016 from http://www.wpro.who.int/publications/docs/EHRmanual.pdf

[18] Tierney, W. M., Achieng, M., Baker, E., Bell, A., Biondich, P., Braitstein, P., ... & Musinguzi, N. (2010). Experience implementing electronic health records in three East African countries. Stud Health Technol Inform, 160(Pt 1), 371-5.

[19] Bahamdain, S. S. (2015). Open Source Software (OSS) Quality Assurance: A Survey Paper. *Procedia Computer Science*, *56*, 459-464.

[20] Khanjani, A., & Sulaiman, R. (2011, March). The process of quality assurance under open source software development. In Computers & Informatics (ISCI), 2011 IEEE Symposium on (pp. 548-552). IEEE.

[21] Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012, February). Social coding in GitHub: transparency and collaboration in an open software repository. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (pp. 1277-1286). ACM.

[22] Carey, G. E., & Braunack-Mayer, A. J. (2009). Exploring the effects of government funding on community-based organizations:'top-down'or 'bottom-up'approaches to health promotion?. Global health promotion, 16(3), 45-52.

[1] http://www.fsf.org/
[2] https://opensource.org/osd
[3] https://github.com
[4] https://bitbucket.org
[5] https://www.sourcetreeapp.com
[6] https://sourceforge.net
[7] https://code.google.com

Evaluation of PID Controller Performance in Mobile Telemedicine Vehicles Network Within Nigeria

Ajiboye Aye Taiwo
Department of Computer Engineering,
University of Ilorin,
Ilorin, Nigeria
atajiboye@yahoo.co.uk, ajiboye.at@unilrin.edu.ng

Ibiyemi Tunji Samuel
Department of Electrical and Electronics Engineering,
University of Ilorin,
Ilorin, Nigeria.
ibiyemits@yahoo.com, tibiyemi@unilorin.edu.ng

***Abstract* - The availability, quality and cost of healthcare delivery in developing countries in general and Nigeria in particular could be improved upon if information technology is appropriately introduced into the health care delivery scheme and well managed. Therefore, introduction of telemedicine into the developing nation's healthcare delivery scheme could be a solution to a number of challenges facing the scheme including acute shortage of medical personnel. The aim of this work is to have a system that will enable the dishes on the mobile vehicles at the remote location to send/receive correct medical information to/from the base station via Nigcomsat-1R. To achieve this aim there is the need for precise tracking of the satellite by the mobile dishes. Since these vehicles are distributed all over Nigeria, the time delay between the dishes on these vehicles and the base station is enormous, because of Nigeria large land mass of about 923,677 sq km. The effectiveness of Proportional-Integral-Derivative (PID) controller used in this work is largely determined by the amount of delay between remote nodes and the base station as the performance of PID controller degrades with increase in time delay. The minimum and maximum time delay between University of Ilorin and the telemedicine vehicle being at any location within Nigeria was determined. The performance of both the uncompensated and the compensated system was determined and evaluated based on time domain performance indices. The analysis of the simulation results confirmed remarkable improvement in the performance of the compensated system over the uncompensated system.**

Keywords - Mobile telemedicine vehicle; NigComSat-1R; PID controller; time delay; performance indices;

VI. INTRODUCTION

Telemedicine is defined as the provision of health care delivery and sharing of medical knowledge over a distance using Information and Communication Technology (ICT). According to [1], Telemedicine is one of the pilot projects of NIGCOMSAT-1 with an objective to develop a telemedicine system to improve remote diagnosis and to deliver cost effective and qualitative healthcare in Nigeria. The major objective of mobile telemedicine is to improve the quality and timeliness of health care delivery offered during the golden window of treatment opportunity immediately following injury, and to provide better information to health personnel just before physical presence of the patient at the health centre [2]. Integration of Information and Communication Technology, ICT, into health care delivery is making it more affordable and available even in rural areas without sacrificing the quality of delivery as obtainable in urban cities [3]. Investing in the development of Telemedicine technology if properly managed can serve as a way to alleviate the effects of acute shortage of medical personnel and the low level of healthcare delivery in developing countries in general and Nigeria in particular.

The link between the base station and the distributed mobile stations is via Nigerian's geostationary communication satellite (Nigcomsat-1R) which is located in ITU's region1 at 42.5^{o} East and distance of 35768 km from the Equator.

The aim of this work is to have a system that will enable the dishes mounted on the mobile vehicles to send/receive correct medical information to/from the base station via Nigcomsat-1R.

The required command and control signals for the dishes are to come from the base station. To achieve the required connection between the base station and these distributed mobile stations there is the need for precise tracking of the geostationary satellite (Nigcomsat-1R) by satellite dish mounted on the mobile station.

The proposed controller for the control and command is Proportional-Integral-Derivative (PID) controller algorithm. PID controller is the most form of feedback in used in industries [4]; more than 90% of all control loops are PID. Therefore, PID is the first solution that should be tried when feedback is used [5].

Since the system we are dealing with involved mobile nodes which are highly distributed in nature the expected time delay will be large and variable due to large Nigeria land mass and node mobility respectively. Hence the effectiveness of PID controller algorithm in this type of system will be reduced because it depends on the amount of delay between formulation of PID control law and its delivery to the actuator [3], [6]. The situation becomes worse with increase in time delay.

The major factor that determines the magnitude of the time delay is the position within Nigeria of a telemedicine

vehicle with a satellite dish mounted on it with respect to the base station. Because of large land mass of about 923,677 sq km [3], [6] the time delay incurred can be very large.

In this research work, time delay between the base station and the distributed nodes was modelled and determined using simulation. Based on the determined delay the effectiveness of already designed PID controller by [7] for mobile satellite dish within Nigeria was assessed using time domain performance indices.

VII. DETERMINATION OF TIME DELAY RANGE

The geographical coordinates of Nigeria is between longitude 2°43.207'E and 14°54.685'E and latitude 4°17.825'N and 13°52.837'N. It is assumed that the Base station is at University of Ilorin which is located at 4°40.500'E, 8°29.100'N and the dish mounted telemedicine vehicles distributed all over Nigeria as shown in Fig. 1. The location and geographical coordinates of points A, B, C and D presented in Fig. 1 are as follows: point A is located at Niger Republic and its coordinate is 2^{o}43.207'E, 13^{o}52.837'N; point B is located at Lake Chard and its coordinate is 14°54.685'E, 13°52.837'N; point C is located at Gulf of Guinea and its coordinate is 2°43.207'E, 4°17.825'N and finally point D is located at Cameroon and its coordinate is 14^{o}54.685'E, 4^{o}17.825'N.

The subsatellite point for NigComSat-1R satellite is at 42.5^{0}E, 0^{0}.

The parameters needed for the determination of time delay is the distance between the base station and mobile telemedicine vehicles. The equation for the determination of the distance between two points on the earth surface via any geostationary satellite is as given in (1) [3], [8], [9].

Fig.1. Nigeria Geographical Map showing distribution of Telemedicine Vehicles

Therefore, the time delay between these two points can be determined by dividing this distance by the signal speed which is assumed to be the speed of light (30 * 10^{9} m/s.).

$$d_{sr} = \sqrt{D^2 + R^2 - 2DR\cos(\alpha_{sn})\cos(\Delta_1)} + \sqrt{D^2 + R^2 - 2DR\cos(\alpha_{rn})\cos(\Delta_2)}$$

Where:

$$\Delta_2 = \Delta_{rn} - \Delta_S$$

= distance between the source and the receiving node

= radius of the earth in km

= sum of the radius of the earth and satellite altitude in km

= angle of longitude of the subsatellite point in degrees.

= latitude of the sending node location on the earth surface in degrees.

= latitude of the receiving node location on the earth surface in degrees.

= angle of longitude of the sending node location on the earth surface in degrees.

= angle of longitude of the receiving node location on the earth surface in degrees

= signal speed in m/s

The simulation of the time delay was carried out in MATLAB environment and the resulting simulation graph is as shown in Fig. 2. From the Simulation graph of Fig. 2 the maximum and minimum time delay was determined to be 0.2502 second and 0.2469 second and they occurred at point A and point B as shown on the graph of Fig. 2. Points A and B are at Niger Republic (2^{0}43.207'E, 13°52.837'N) and Cameroon (14^{0}54.685'E, 4^{0}17.825'N) respectively.

Point A on the graph of Fig. 2 corresponds to point A in Fig. 1 and point B on the graph of Fig. 2 corresponds to point D in Fig. 1.

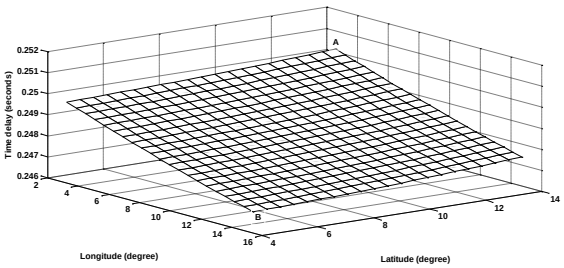


Fig.2. Time delay variations between the base station at Ilorin (4°40.500'E, 8°29.100'N) and any telemedicine vehicle with a dish mounted located within Nigeria and communicating via Nigcomsat-1R

VIII. DETERMINATION OF SYSTEM TRANSFER FUNCTION

The transfer function was determined from the dish structure, jack actuator and time delay dynamics model. The determination of the dish dynamics was based on the moment of inertia of the dish structure (satellite dish and BUC/LNB), spring constant, and damping coefficient.
These parameters were determined by experiment because their value cannot be read off the plant at the node [9].
Due to the same reason the actuator jack dynamics was also empirically determined.
The resulting closed loop distributed control system can be represented by the block diagram of Fig. 3. According to [3], [6], [7], [8] the transfer function for this system is as shown in (2).

$$\frac{\theta_A(s)}{\theta_r(s)} = \frac{G_c(s)G_p(s)G_{d1}(s)}{1+G_c(s)G_p(s)G_{d1}(s)G_{d2}(s)} \quad (2)$$

Where:
$\theta_A(s)$ = Actual dish position in degree;
= reference dish position in degree;
$G_c(s)$ = PID controller transfer function;
$G_{d1}(s)$ = Forward path time-delay transfer function;
$G_p(s)$ = plant transfer function
$G_{d2}(s)$= Feedback time delay open-loop transfer function;

Transfer function for PID controller is as defined in (3):

$$G_c(s) = \frac{K_d s^2 + K_p s + K_i}{s} \quad (3)$$

Where:
= Proportional gain;
= Integral gain;
= Derivative gain

The plant, transfer function is made up of the transfer function of the dish structure, and actuator jack and is as expressed in (4):

$$G_p(s) = \frac{3.76}{s^4 + 67.56\,s^3 + 62.36\,s^2 + 150.52s} \quad (4)$$

The transfer function for the Forward path and Feedback time delay is as expressed in (5).

$$\left.\begin{aligned} G_{d1}(s) &= e^{-T_1 s} \\ G_{d2}(s) &= e^{-T_2 s} \end{aligned}\right\} \quad (5)$$

Where:
= Time delay between source and receiving node (Forward path time-delay)
= Time delay between receiving and source node (Feedback time delay)

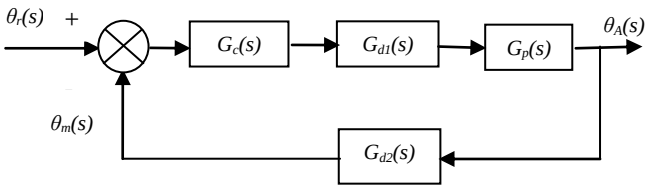


Fig.3. Block diagram of the Composite System

Substituting (3), (4) and (5) in (2) gives the system closed loop transfer function shown in (6).

$$\frac{3.76\left(K_d s^2 + K_P s + K_i\right)e^{-T_1 s}}{s^5 + 67.56s^4 + 62.36s^3 + 150.52s^2 \ldots . + 3.76(K_d s^2 + K_P s + K_i)e^{-(T_1+T_2)s}} \quad (6)$$

The effectiveness of the PID controller in this system was evaluated by comparing the time domain performance indices that is the rise time, (, Peak time ( ), settling time ( ) and percentage overshoot ( ) for the uncompensated system with that of compensated system when the system is subjected to a unit step function. The performance indices for the uncompensated system form the basis for the determination of effectiveness of the PID controller in the improvement of system performance. The maximum obtainable time delay (0.2502 seconds) being the worse case was used for the simulation.

IX. SIMULATION AND ANALYSIS

Simulation experiment was carried out in MATLAB environment to evaluate the degree of improvement in system performance indices resulting from the incorporation of PID controller into the system. The PID parameter values obtained by [7] for optimisation of system settling time for mobile satellite dish position control within Nigeria (K_p=20, K_i=4 and K_d=0) was used for the simulation because these controller gain values were determined for similar system within Nigeria terrain.
The time domain performance indices obtained from the system step response of Fig. 4 for both the uncompensated and compensated systems are as presented in Table 1.

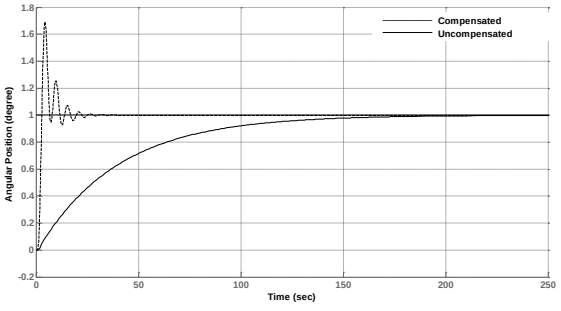


Fig.4. System step response for $K_p = 20$, $K_i = 4$ and $K_d = 0$

TABLE I. SYSTEM TIME DOMAIN PERFORMANCE INDICES

System condition	Percentage overshoot (%)	Settling time (seconds)	Rise time (seconds)	Steady state error
Uncompensated	0	154	85.9	0 at 250 seconds
Compensated	69	21.4	1.3	0 at 30 seconds

From Table 1 the time required for the uncompensated system to rise, settle and achieve zero steady state error is enormous compare to that of compensated system.
The only short coming noticed as a result of incorporation of PID controller is the overshoot, but this effect reduced from 69% at 4.5 seconds to: 25% at 10 seconds, 7% at 15 seconds and finally 0% at 30 seconds. This result shows an improvement in the performance of the PID controller compensated system over the uncompensated system.

X. CONCLUSION

Time delay model for the determination of the delay between any two points on the earth surface via geostationary satellite was derived. The minimum and maximum time delay between University of Ilorin and any location within Nigeria via Nigcosat-1R was determined The performance of PID controller in the tracking of geostationary satellite (Nigcomsat-1R) by dish mounted on mobile telemedicine vehicles within Nigeria was evaluated based on time domain performance indices. The results from the simulation experiments shows remarkable improvement in the performance index of the compensated system over the uncompensated system.

References

[1] N. C. Anyaegbunam Felix "On National e-Healthcare Delivery through Nigcomsat-1," International Journal of Engineering Research & Technology, Vol. 3, Issue 1, pp. 2270 – 2276, January – 2014.

[2] J. S. Cullen, D. M. Gagliano, W. R. Gaasch, R. D.P. Gunawardane, P. F-M. Hu, M. P. LaMonte, C. F. Mackenzie and Y. Xiao "Mobile Telemedicine Testbed," Systems and Information Technology Review Journal, pp. 1- 16, Spring/Summer 1999.

[3] T. S. Ibiyemi, and A. T. Ajiboye. "Automatic Tracking of NigComSat-1R Satellite by Dish Network Mounted on Mobile Tele-Medicine Vehicles," International Journal of Engineering Research & Technology, Published by International Research Publication House, India. Vol. **1,** Issue **4, pp. 1 – 4, June 2012.**

[4] A. Salem Farhan, "New Efficient Model-Based PID Design Method," European Scientific Journal Edition, Vol. 9, No. 15, pp. 181-199, May 2013.

[5] K.J Astrom, T. Hagglund, "The future of PID control," Control Engineering Practice 9, pp. 1163–1175, 2001.

[6] A.T. Ajiboye, "Development of Automatic Position Controller for Mobile Satellite Dish Network within Nigeria," Ph.D. Thesis, Department of Electrical Engineering, University of Ilorin, Ilorin, Nigeria, July 2012.

[7] A. T. Ajiboye, T. S. Ibiyemi and J. A. Falade "Optimisation of Settling Time for PID Position Control in Mobile Satellite Dish Network within Nigeria," *Journal of Telecommunications*, Vol. 20, Issue 2, pp.22–26, June 2013.

[8] T. S. Ibiyemi, and A. T. Ajiboye. "On PID Controller Design for Mobile Node Network with Variable Delays," International Journal of Science and Advanced Technology, Publisher by Dhaka: International journal of Science and Advanced Technology, Bangladesh. Vol. 2, No. 4, pp. 176–180, 2012.

[9] T. S. Ibiyemi, and A. T. Ajiboye, "Robust PID Control of Mobile Satellite Dish Network within Nigeria," International Journal of Computer Applications, Vol. 41, No. 21, pp.32–36, 2012.

Learner infrastructural capability: A way forward to achieving eLearning success

Naeem Atanda Balogun
[1]Department of Information and Communication Science
University of Ilorin
Ilorin, Nigeria
neembaloo@yahoo.com

Abd Rahman Ahlan
Department of Information Systems International Islamic University Malaysia Gombak, Malaysia
aahlan@gmail.com

Abstract **- This study introduces learner infrastructural capability into the field of information systems success. The introduction of this construct in information systems success is important due to lack of consideration for the individual infrastructural ability that has not been addressed by the success factors with respect to individual, environment, developing countries and digital divide. An explanatory mixed method research design is considered using questionnaire for the survey and an open-ended interview for data collection. 412 responses were used from survey and 10 people for interviewed. The study found a good model fit when learner infrastructural capabilities is merge within the information systems success model of DeLone and McLean forming a success model that inculcate the individual IS infrastructure.**

Keywords - infrastructural capability, eLearning, IS, success model

I. INTRODUCTION

Delone and McLean did a tremendous work by developing an Information Systems success model that witnesses a wide usage and acceptance by many information systems researchers. This indicates its success as a success model within the field of information systems however, suggestions by [1] to involve the technology and individual traits were less considered by the researchers. This study tends to involve the individuality, environmental and technological aspect into the Information systems by considering only the end-user in eLearning, i.e. the students.

Learner Infrastructural capabilities which can be considered to be "fit with various information systems activities" by [2] in this study refers to learners' capabilities to personally access the eLearning system in the developing countries. While developing countries are faced with different infrastructural problems in order to have access to learning facilities, developed countries are faced with digital divide. Therefore, consideration should be given to the end-users for personally having the ability to learn without facing any problem due to the environment and technology. In a survey of eLearning in Africa, [3] suggest access to computer, reliable electricity and faster connectivity perhaps might makes eLearning effective based on the response from the survey.

Having these infrastructures in learning institutions within Africa seems to be a dream which is yet to see the light of the day when compared with other continents, and was highlighted to be issues facing eLearning in developing countries by [4]. Perhaps, personally having these infrastructures by learners might help to achieve their eLearning goal. This study therefore, identified these infrastructures as factors and then grouped together as a construct and are refers to as "Learner Infrastructural Capabilities," and these factors are: energy generation, computer ownership, and internet access.

II. INFORMATION SYSTEMS SUCCESS MODEL OF DELONE AND MCLEAN

Information Systems Success model begins with the work of [1], it plays an important position in determining the success of an IS either in an organisation or individually. DeLone and McLean describe their work using the work of Shannon and Weaver (1949) and that of Mason (1978) that describes the concept of level of output from communication theory with the linear nature of information. This inform the Information systems initiation of information then communicates to the recipient who is then influenced by the information, showing information from production through the use has an influence on individual or organisation or both.

When evaluating the success of the system, different researchers in Information Systems try to postulate the relationship between the "organizational performance" and different factors of IS success [1]; [5]; [6]; [7]; [8]. However, based on successful review of nearly Two hundred papers in 1980s, [1] created taxonomy of Information systems and successfully developed an information systems success model by identifying six

important components of information systems success. These are shown in figure 2.1 and include system quality, information quality, use, user satisfaction, individual impact and organizational impact.

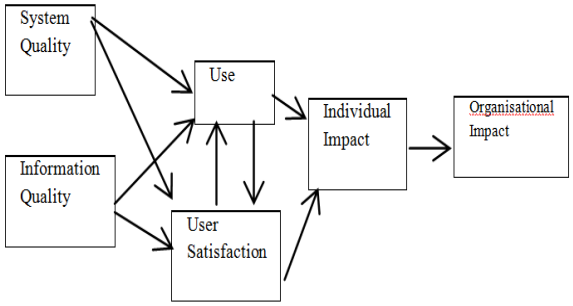


Figure 1: DeLone and McLean IS success model [1].

DeLone and McLean call for further development and validation of their model after the publication in 1992. According to [6] in their evaluation of the information systems success model using four construct i.e. system quality, information quality, use and user satisfaction, modified the construct "use" with "usefulness" and argued that for mandatory system, "usefulness" is a better measure of information systems than "use", while a positive significant relationship was found between the four variables of the study. The following year, [7] argued that IS use is a behavior and it is separated from the IS success model and claimed that systems failed because they provide no benefit and not because they were not used. Meanwhile in [9] argued that the construct "use" needs to be maintained that even in mandatory systems, there can still be considerable variability of "use" and then developed a modified model in figure 2.

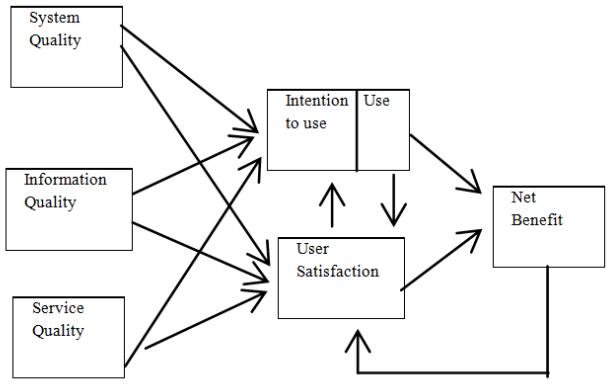


Figure 2: DeLone and McLean IS success model [9].

In 2003, DeLone and McLean further claim that their original model exceeded their expectation, after searching for papers that referenced their model, the result is close-to 300 refereed papers in journals and proceedings just within 1993 to mid-2002. Some researchers tested and validated IS success model [10]; [11], relationship among the variables [12]; [13], while some researchers integrated the variable with Technology Acceptance Model [14], in E-learning [15]; [11] in Leaning Management Systems [16], some researcher in knowledge management [17]; [18] and [19] using multidimensional success measurement instrument provide higher content validity.

These therefore, give the model strength and make it an important model in the field of information systems however, the need to consider other factors had aroused when looked carefully in the issue of information systems in the developing countries and the case of digital divide in the developed nations.

III. LEARNER INFRASTRUCTURAL CAPABILITY

Learner infrastructural capability construct, which comprises of computer ownership, energy generation and internet access within the context of eLearning simply means; learners having personal computer for study. This computer might be a laptop, desktop, palmtop computer, smart mobile phone or personal digital assistant [4]. Having computer for educational purpose is identified to boost students' academic performance [20]; [21]. Technological diffusion in a society could motivate individual to increase their ability to personally own a computer [22], and almost two-third of the students that respond to a questionnaire in a Ghanaian University are found to own a computer [23]. This shows the extent to which the technology is been accepted, use in Africa and the will to having personal computer by students when it is not considered compulsory. Tangoe, found a relationship between computer ownership and frequency of internet usage. According to [24] in their study, found a significant positive association between computer ownership and higher performance in examination. These indicate that personal computer could perhaps have some relationship with internet access and using the computer will increase the students' performance academically and relate socially.

This study refers to internet access as the accessibility of learner to the web and other related systems that connect through the personal computer. According to [20] not having access to internet by students in their households due to poverty results to poor performance at school. While [25] found students to face problem of internet access in the developing countries, and could cause low performance, [26] highlighted this low penetration and low access to internet to be due to developing countries' regulation. Motivation to having a computer and get internet access increases fast due to technological spread [22]. Learners can search for information online with the aide of the internet and adapt to online learning environment that shows flexibility in terms of learning methodology and pedagogy approaches [25]. These also indicate a relationship between having a computer and internet access, as it becomes difficult for an individual to have internet access without having a personal computer as defined by

this study. The study shows that access to internet may assist learners in their successful study and acquiring more knowledge by staying active online and lack of access to internet might cause less active to an information systems.

Electricity generation is making electricity available by individual through the government grid or by an alternative means such as batteries, solar power, wind power and power bank. [3] identified non-availability of electricity as one of the problem faced when delivering eLearning in Africa. The electricity plays a great role in the information systems field, without electricity, all the infrastructures is down and nonfunctional. This makes electricity an important aspect of information system which a less attention is paid to [25] also claimed that electricity is one of the factors disrupting eLearning in Tanzania. According to [27] end-user communication and computing equipment such as personal computers, local servers, wireless routers, set-top, switches, computer monitors and smart TVs account for a total of 55 percent of electricity usage, data centres accounting for almost 30 percent of electricity usage. This shows the capacity of electricity needed by the end-user and if the information systems should be successful it requires a great capacity of electricity and for the developing countries where electricity is less or epileptic, information systems is less active and problems encountered. However, power interruption by national grid might be an issue for information systems success in developing countries, individuals can get alternative means of power source such as using power generator set, solar energy and power bank because the end-user needs more electricity supply.

IV. METHODS

The study employs an explanatory mixed methods design by using a questionnaire for the survey and interview as the second part. The study uses all the 412 responses and a one-on-one interview were conducted for 10 individuals. The responses from survey were obtained from students who enrolled for eLearning within Nigerian Universities, while 6 interview were among students, 4 were conducted to IT professionals.

The study review literature to formulate three dimensions for the constructs in which each dimension is having four to five items, making a total of 13 items for the construct. A content analysis was carried out among senior academic staff, pilot study was conducted among students and the construct having a Cronbach's Alpha of 0.878. After preparing data following thoroughly the steps by [28] analysis were conducted using both IBM SPSS and AMOS (version 21.0). Exploratory factor analysis (EFA) and reliability test were performed to check the reliability and validity of the construct and found to be satisfactory. And confirmatory factor analysis (CFA) was used for validity and assessing the fitness of the measurement model of the construct, and Structural Equation Model (SEM) for testing the structural model fit.

Descriptive Statistics was conducted to identify missing values however, no missing value was found. Correlation Matrix to identify singularity issues, by checking items that has correlation of more than 0.2 with other item and determinant greater than 0.00001, no singularity issue identified. Kaiser-Meyer-Olkin (KMO) is 0.922 and Bartlett's test with significant value of 0.000 showing sample adequacy and no identity matrix found. Anti-image matrix was checked, all values are greater than 0.5 showing satisfaction. Communalities shows no item less than 0.5 showing satisfaction and Total variance explained was satisfactory and no cumulative percentage is greater than 90 percent, showing satisfaction. Reliability was conducted on the construct, all items are found to be satisfactory with the lowest and highest having Cronbach's Alpha of 0.914 and 0.917 respectively. This shows that the items are reliable and valid.

After a successful exploratory factor analysis (EFA) and the check for reliability of the factors and their items, then Confirmatory Factor Analysis (CFA) is used according to [29] to test the reliability of the observed variables, examine interrelationships and correlation among the latent construct, to find the best indicators for latent variables before the full fledge Model. CFA was performed on the construct to find the observed variables that best described the latent variable and the interrelationship and correlation among the factors.

In [30] suggestion was made to report Chi-Square Test (X^2), degree of freedom (df), p-value, Root Mean Square Error of Approximation (RMSEA), Comparative Fix Index (CFI) and parsimony fit index such as the PNFI having cut-offs of RMSEA < 0.1 is acceptable and best < 0.08, CFI > 0.9 is good while > 0.95 is best, Normed Chi-Square (X^2/df) < 5. Based on the three dimensions used to measure the leaner infrastructure capability, variables showing the factor correlation, factor loading, error variance are assessed. The model fits results are shown, indicating a good model fit to data (X^2 = 201.372, df = 62, p = 0.000, X^2/df = 3.248, CFI = 0.947, RMSEA = 0.074). Likewise, in a second order, the result also show a model fit without difference (X^2 = 201.372, df = 62, p = 0.000, X^2/df = 3.248, CFI = 0.947, RMSEA = 0.074). *See* Figure 3 and 4.

Considering the structural equation model, having the dimensions of the Information systems success models and the Learner infrastructural capability, the model is found to be fit using the suggestion by [30]. The IS success model was found to be fit in the study, having (X^2 = 855.613, df = 310, p = 0.000, X^2/df =2.760, CFI = 0.932, RMSEA = 0.065). and the model that introduced Learner infrastructural capabilities into the Information systems success model was also found to be fit, having the following readings: (X^2 = 1736.047, df = 721, p = 0.000, X^2/df =2.408, CFI = 0.909, RMSEA = 0.059).

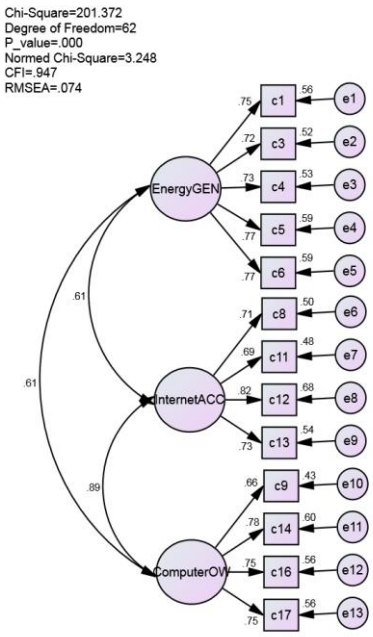


Fig 3: CFA of LIC

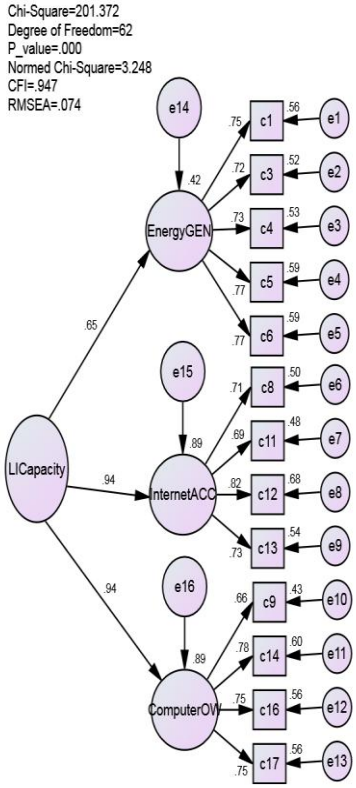


Figure 4: CFA of LIC

V. CONCLUSION

Upon the success shown by the IS success model in the field of Information systems, a consideration for individual, environment in developing countries, and the digital divide which had been an issue within the developed countries is highlighted. This study thereby, show that, to achieve success in information systems, tangible consideration need be given to the ability of the individual to personally have the infrastructure it takes to meet up with the technological challenges. Having the "learner infrastructural capability" which include computer ownership, internet access and energy generation will go a long way in determining the success of information systems in eLearning.

REFERENCES

[1] W.H. DeLone, and E.R. McLean, "Information systems success: the quest for the dependent variable," Information systems research, vol. 3, pp. 60-95, 1992.

[2] P. Weill, & M.H. Olson, "An assessment of the contingency theory of MIS," Centre for Digital Economy Research, Stern School of Business, 1987.

[3] T. Unwin, "Survey of e-Learning in Africa: based on a questionnaire survey of people on the e-Learning Africa database in 2007," Royal Holloway, University of London. 2008.

[4] N.A. Balogun and A. Ahlan, "Issues of e-learning in developing countries: a Nigerian perspective," IEEE Xplore ICT4M. Malaysia, pp. 1-4, November 2014.

[5] P.B. Seddon and M.Y. Kiew, "A partial test and development of the DeLone and McLean model of IS success," Association for Information Systems, pp. 99–110, 1994.

[6] P.B. Seddon and M.Y. Kiew, "A partial test and development of DeLone and McLean's model of IS success," Australian Journal of Information Systems, vol. 4, pp. 90–109, 1996.

[7] P.B. Seddon, "A respecification and extension of the DeLone and McLean model of IS success," Information Systems Research, vol 8, pp. 240–253, 1997.

[8] P.B. Seddon, D.S. Staples, R. Patnayakuni, and M.J. Bowtell, "The dimensions of information systems success," Communications of the Association for Information Systems, vol. 2, pp. 1-13, 1999.
[9] W.H. Delone, & E.R. McLean, "The delone and mclean model of information systems success: a ten year update," Journal of management information systems, vol. 19, pp. 9-30, 2003.
[10] A. Rai, S.S. Lang, and R.B. Welker, "Assessing the validity of IS success models: An empirical test and theoretical analysis," Information Systems Research, vol. 13, pp. 50–69, 2002.
[11] R.D. Freeze, K.A. Alshare, P.L. Lane, and H.J. Wen, "IS success model in e-learning context based on students' perceptions," Journal of information systems education, vol. 21, pp. 173-184, 2010.
[12] M.A. Mahmood, L. Hall, & D.L. Swanberg, "Factors affecting information technology usage: a meta-analysis of the empirical literature," Journal of Organizational Computing & Electronic Commerce, vol. 11, pp. 107–130, 2001.
[13] R.H. Bokhari, "The relationship between system usage and user satisfaction: a meta-analysis," The Journal of Enterprise Information Management, vol. 18, no. 2, pp. 211–234, 2005.
[14] B.H. Wixom, and P.A. Todd, "A theoretical integration of user satisfaction and technology acceptance," Information Systems Research, vol. 16, pp. 85–102, 2005.
[15] A. Lee-Post, "E-learning success model: an information systems perspective," Electronic journal of e-learning, vol. 7, pp. 61-70, 2009.
[16] X. Song, "Teaching and learning experience with learning management systems: an adapted IS success model in LMS context," Proceedings of the International symposium on IT in medicine and education (ITME), pp. 148-152, 2011.
[17] M.E. Jennex and L. Olfman, "Organizational memory/knowledge effects on productivity: a longitudinal study," J.R. Sprague, Eds, pp. 109, 2002.
[18] J.H. Wu, & Y.M. Wang, "Measuring KMS success: a specification of the DeLone and McLean model," Information & Management, vol. 43, pp. 728–739, 2006.
[19] D. Sedera, G. Gable, and T. Chan, "A factor and structural equation analysis of the enterprise systems success measurement model," Association for Information Systems, Washington, DC, pp. 448-464, 2004.
[20] J. Burns, "A third of poorest pupils without internet at home," BBC news education and family, 2013.
[21] A. Ahlan, N.A. Balogun, and Y.I. Shehu, "Information technology enhances students' academic performance: A case of University of Ilorin," The Online Journal of Distance Education and e-Learning, vol. 2, pp.15-20, 2014.
[22] J.A. Van DIJK, The evolution of the digital divide-the digital divide turns to inequality of skills and usage, IOS Press, Digital enlightenment yearbook, 2012.
[23] M. Tagoe, "Students' perceptions on incorporating e-learning into teaching and learning at the University of Ghana," International Journal of Education and Development using Information and Communication Technology, vol. 8, pp. 91-103, 2012.
[24] J. Schmitt, & J. Wadsworth, "Is there an impact of household computer ownership on children's educational attainment in Britain?" Centre for economic performance. Discussion paper no 625, 2004.
[25] V. Ndume, F.N. Tilya, and H. Twaakyondo, "Challenges of Adaptive eLearning at Higher Learning Institutions: A Case Study in Tanzania," International Journal of Computing and ICT Research, vol. 2, pp. 47 – 59, 2008.
[26] S. Wallsten, Regulation and Internet Use in Developing Countries. World Bank Policy Research Working Paper No. 2979, 2002.
[27] J. Koomey, "Energy use and the information economy, 2014.
[28] F.J. Jr. Fowler, Survey research methods, 2nd ed., Newbury: Sage Publications, 1993.
[29] J.B. Schreiber, F.K. Stage, J. King, A. Nora, and A.E. Barlow, "Reporting Structural equation Modeling and Confirmatory Factor Analysis Results: A Review," The Journal of Educational Research vol. 99, pp. 323-337, 2006.
[30] D. Hooper, J. Coughlan, & M.R. Mullen, "Structural Equation Modelling: Guidelines for Determining Model Fit," The Electronic Journal of Business Research Methods. Vol. 6, pp. 53 – 60, 2008.

Proceedings of the 1St International Conference of IEEE Nigeria Computer Chapter In collaboration with Department of Computer Science, University of Ilorin, Ilorin, Nigeria - 2016

Factors Determining Effectiveness of E-Learning in Secondary Schools

[1]Subramaniam Palanisamy
[2]Abd Rahman Ahlan
[1, 2]Department of Information Systems International Islamic University Malaysia Gombak, Malaysia
[1]palasubra@hotmail.com, [2]aahlan@gmail.com,

Naeem Atanda Balogun[3]
[3]Department of Information and Communication Science
University of Ilorin,
Ilorin, Nigeria
[3]neembaloo@yahoo.com

***Abstract* – This study explore factors to determine e-learning effectiveness in secondary school within Malaysia. The study used Technology Acceptance Model, Information Systems Success Model to propose a Model that can be used to determine effective use of e-learning in secondary school. The study is targeted at only secondary school students.**

Keywords – Malaysia secondary school, e-learning, effectiveness, information systems

I. INTRODUCTION

Education is increasingly embracing active learning model over the traditional transmission of transaction. Advanced technology has led to the conversion of the traditional courses into electronic learning (e-learning). Essentially, e-learning unlike traditional learning has become another way of teaching and learning. Learning has shifted from something that is done at a specific time, in a specific place, for a limited number of individuals to a resource that is available virtually anytime, anyplace, to anyone [1]. Therefore, e-learning is considered as an educational paradigm [2] and it uses of internet technology to enhance the user's knowledge and performance. In fact, Information Technology in teaching and learning has necessitate to transform how students learn by using more advanced, efficient, and effective alternatives such as e-learning [3]. In addition, it also involves the delivery of course content via electronic media, such as Internet, Intranets, Extranets, audio or video tape, satellite broadcast, interactive Television and CD-ROM [4]. Also, it has been viewed as synonymous with Internet-Based Training (IBT), Web-Based Learning (WBL), Web-Based Instruction (WBI), Advanced Distributed Learning (ADL), Online Learning (OL) and Open or Flexible Learning (OFL) [5].

The rapid growth of e-learning in education has dramatically increased and attracted much attention of learners and the service providers. There are various reasons which motivated the use of e-learning in Malaysian schools at all levels by government and private schools, these include, e-learning proven to be effective way of learning in schools especially in developed countries, the introduction of e-learning in kindergartens, children at young age were exposed to the use of electronic for learning in early education, the introduction of e-learning in schools and higher learning institutions with synchronous and asynchronous mode of teaching and learning, the need for students to spend more time on learning by going for self-pace learning and self-directed learning. Therefore, e-learning has become a generally acceptable option of learning practice in all type of learning institutions. The importance of e-learning as a new mode of learning and it is in the process of succeeding the traditional face-to face learning is widely acknowledged. This significance sparks the needs for the effective implementation of e-learning in learning institution such as higher learning institutions and schools especially in developed and developing countries. This has prompted educational institutions to focus more on e-learning and striving to make it successful.

The effectiveness of e-learning in schools need to be research into as the technology gains advancement in providing better computer mediated instruction, electronic communication and content delivery. Scholars are conducting research on the factors that determine the effectiveness of e-learning among e-leaning students in developed countries. The majority of e-learning research in secondary schools is focused on the e-learning students' intrinsic motivation to succeed in independent learning environments [6]; [7]. A study on e-learning students on technology and isolated learning environment revealed that they were struggling with e-learning technology and isolated learning environment but enjoyed autonomy provided by the e-learning online instruction [8] and [9].

Scholars focused more on e-learners in developed countries but e-learning in developing countries still need research attention. In Malaysia, many higher learning institutions adopt e-learning systems for their e-learning courses. However, the adoption of e-learning in Malaysian schools, notably is not wide because the government only implemented e-learning in selected schools at the beginning. At first, e-learning was introduced in smart schools in 1999 and followed by introduction of Learning Management System (LMS) in 120 selected schools. Later, in 2012, Virtual Learning Environment (VLE) FROG was introduced in 10 000 schools in Malaysia.

Notably, the implementation of e-learning in Higher Learning Institution is different than in Malaysian secondary

schools in terms of teaching and learning, assignments and evaluation. The students in Higher Learning Institution deploy full e-learning mode unlike students in secondary schools, who use blended system, partially using e-learning system and undergoing face-to-face teaching and learning with their teachers in the classroom. The students in the schools need more attention and assistance from their teachers even though they can access the e-learning system, e-learning materials, answer the given questions or assignments using online system similar to the students in Higher Learning Institution.

Even though e-learning initiative in Malaysian schools have been in practice for almost a couple of decade, there has not been any research done on the e-learning students' perspective to find the problems faced by students to learn effectively using the system. Even then, there were some research conducted on the impact of the teaching and learning on e-learners using VLE FROG Platform by EPRD. The research was focused on students knowledge, skill, motivation and attitude. The general finding shows that 636 respondents (31.3%) lack ICT skills and 180 respondents (8.8%) were having insufficient knowledge of using e-learning system. Whereas, their motivation level and attitude towards using e-learning system is at high level.

The main problems faced by the students in schools are difficulties in accessing the system, slow internet, lack of facilities and lack of training to use the system well. These problems contributed much to the ineffectiveness of e-learning. Subsequently, jeopardize the whole objective of implementing e-learning system in Malaysian schools.

In 2001, [10] discovered that the students became active participants in their own learning, and technology assist them in their construction of knowledge. The achievement level of the students has been extensively studied by the researchers from west especially in Higher Learning Institutions [3]. According to [10], Director, Center for IT in University Malaya stated that, even though there are many research conducted on e-learning, most of the research focused in higher learning institutions. There were very few researches conducted on students in secondary schools. This is because greater implementation of e-learning has been noted at the higher education level whilst still in its infancy at the school level [11]. Since there were not many research conducted on the factors that determine perceived effectiveness of e-learning in Malaysian secondary schools, the researcher intends to explore and understand the factors that determine the effectiveness of e-learning among students in secondary school and propose a conceptual model which explains the factors that determine the effectiveness of e-learning and provide suggestions to improve their e-learning.

This study focuses on the use, users' satisfaction of the e-learning system and their related factors which determines the effectiveness of e-learning. The antecedent factors for the effectiveness of e-learning are use of e-learning and users' satisfaction. Whereas, perceived ease of use, perceived usefulness, delivery method, content quality, system quality and self-efficacy would be the antecedents of use of e-learning system and users' satisfaction. In addition, the use of e-learning also would be an antecedent for users' satisfaction. The study was conducted in nine Malaysian secondary schools in Federal Territory of Kuala Lumpur and Putra Jaya as the selected schools are champion in using VLE Frog platform (Source : Education Technology Division-Federal Territory of Kuala Lumpur and Putra Jaya) as these students had been using e-learning system for a longer period of time. The focus group for the research problem is the e-learning students from the 5 schools in Federal Territory in Kuala Lumpur, Putrajaya and Selangor. Those schools are SMK Putrajaya Prisint 8(1), SMK Putrajaya Prisint 18(1), SMK Putrajaya Prisint 18(1), SMK Seri Bintang Utara, SMK Kiara Mas dan SMV ERT-Setapak.

II. EFFECTIVE E-LEARNING IMPLEMENTATION

A. E-Learning Definition

E-learning is often referred as electronic learning. It is emerging as the new paradigm of modern education and can be considered as an educational paradigm [10]. The e-learning students do not depend on chalk and talk method which is a traditional method. Their teaching and learning process depends on the electronic devices which provide the materials and tutorials as well. Due to the evolution of technology and its associated fields, researchers are yet to find a common definition and terminology for defining the e-learning [12]; [13]. This scenario made it very difficult for the researchers to come up with comprehensive and meaningful definition on e-learning and contributed for creation of many conflicting definitions about e-learning.

B. Benefit of e-learning

Numerous studies reviewed the importance and benefits for the e-learning in schools [14]. E-learning has numerous benefits, offering the learners useful learning method, providing series of modules, assessment that guide the learners through the learning process. In addition, [15] noted that e-learning management system could combine the online and traditional classroom learning in an e-learning environment. They viewed the e-learning as instructional authoring tool which allows the contents to be delivered through CD_ROMs, LAN, internet or satellite broadcast. It includes wide range of applications such as Computer Based Training (CBT), Electronic Performance Support System (EPSS), Web Based Training (WBT), Virtual Classroom and Digital Collaborations. Besides, it also supports collaborative community where it offers multiple modes of learning to the learners and facilities for the parents and schools to have good communication at their own convenience.

C. Implementation of e-learning in Malaysian Smart Schools

The Malaysian smart school is a learning institution that has been systematically reinvented in terms of teaching and learning and management practices to help children to cope with the information age. The school culture would be transformed from a memory based to an informed, thinking, creative and caring one through leading age technology. The smart school project represents an experiment that the country will ensure is right from the start and will be a valuable lesson that the other countries especially developing countries can learn from. The smart schools initiative has five goals which are focused on the need to develop a skilled work force for the information age and to promote the goals of the National Philosophy of Education. They are to produce a thinking and technology literate workforce, democratic education, increased participation of stakeholders, provide an all-round development of the individual (intellectual, physical, emotional and spiritual) and provide opportunities to enhance individual strengths and abilities. In addition to this, Malaysia intends to transform its' educational system, in line with and in support of the nation's drive to fulfil Vision 2020. This vision calls for sustained, productivity driven growth which will be achievable only with technologically literate, critically thinking work force prepared to participate fully in the global economy of the 21st century.

D. Implementation of Virtual Learning Environment Frog (VLE Frog) in all government schools in Malaysia

1BestariNet is a project initiated by the Malaysian Ministry of Education (MOE) and carried out in partnership with YTL Communications. Under the project, 10,000 primary and secondary public schools in Malaysia equipped with high-speed 4G Internet access and a virtual learning platform, providing high-speed Internet connectivity and access to a world-class Integrated Learning Solution. Under 1BestariNet, schools were equipped with an integrated solution allowing teaching, learning, collaboration, and administrative functions to take place through the Internet-based Virtual Learning Environment (VLE Frog), which can be accessed in school and from anywhere else with an Internet connection (Yes 4G Internet).

E. Factors that determine the effectiveness of e-learning

Scholars identified human factor and design factor as antecedents of learning effectiveness [16]. The human factors are referred to students and instructors or teachers whereas the design factors are referred to technology, course contents, learners control over the learning, interaction between the course contents, instructors and peers. [17] Explains human factors as manipulative factors which are influenced by school culture, attitudes and beliefs and these can be managed and changed according to our need. These factors include students and teachers attitude towards technology, their knowledge about technology and the school policy which become determinants of effective use of e-learning in school. Whereas, the non-manipulative factors such as age of the students and teachers are not influenced by the school factors. [18] stressed six elements of e-learning environment in educational institution for effective instructional design principles. Those are the content, structure, motivation, feedback, interactions and learning strategies.

According to a research conduction on the e-learning scenario regarding e-learning technology in Canada by [19], their finding reveals that when e-learning is implemented on K12 students, the impact of e-learning and technology use was highest in e-learning compare with face-to-face instructional settings. When implemented appropriately, technology tools are beneficial to students' learning, and facilitate the development of higher order thinking skills. Also, e-learning students' manipulation of technology in achieving the goals of education is preferable to teacher manipulation of technology. Moreover, network-based technologies (e.g., Internet, Web-based, CMC) produced a higher impact than straight technology integration in educational settings. Interestingly, among the Pedagogical Uses of Technology, student applications (i.e., students using technology) and communication applications had a higher impact than instructional or informative uses. This means that the student manipulation of technology in achieving educational goals is preferable to teacher manipulation of technology.

E-learning students' self motivation plays important role in making the e-learning to be effective and benefit learners. Self-motivation is defined as the self-generated energy that gives behavior direction towards a particular goal [20]. Besides, self-motivation is the core of self-regulated learning [21]. The self-regulatory attributes are the learner's personal learning characteristics. Learners characteristics, such as self-efficacy, self-directed behavior, and autonomy need to be identified to make them to become active learners and their learning becomes effective [22]. Furthermore, students' engagement with learning activities can improve learning outcomes such as problem solving skill and critical thinking skill [23]. So, since students are the primary participant of the e-learning system, the great concern must be place on them so that they become responsible on the learning, active and manage their learning process well.

Self-efficacy is one of the learner's personal characteristics which refer to a situation-specific self-confidence in one's abilities [24]. [25] cited self-efficacy as one of the important element of learners' personal characteristics which should be considered when facilitating e-learning system. In a nutshell, if the learners have less self-efficacy towards using e-learning system due to dissatisfaction, then it will cause negative impact towards the effectiveness of e-learning.

Course Structure is strongly correlated to user satisfaction and perceived learning outcomes, especially when the course material is organized into logical and understandable components and that the clear communication of course objectives and procedures will lead to the high levels of student satisfaction and perceived learning outcomes [26]. Course structure refers to course expectation and course infrastructure. Course expectation covers the syllabus,

quizzes, tests, assignments and so forth. Course infrastructure is about the usability of the e-learning system. It must lead them to learn and acquire knowledge. In addition, [27] also cited that information delivery technology, course content, instructor support, mentoring, instructor-to-student interaction, student-to-student interaction and course structure are the main contributing factors for the effectiveness of e-learning.

Another extensive study was conducted [28] to examine the relationship among students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in e-learning settings. The results showed e-learning students with previous e- learning experiences tended to have more effective learning strategies and had higher levels of motivation in their e-learning courses. In addition, when students had higher levels of motivation in their e-learning courses, their levels of technology self-efficacy and course satisfaction increased. Finally, students with higher levels of technology, self-efficacy and course satisfaction earned better final grades.

II. MODEL FOR EFFECTIVENESS OF E-LEARNING

E-learning is viewed as several human and non-human entities interacting and indulging in learning process together through computer and internet or intranet based instructional system whereby the students' satisfaction and use of e-learning system are widely cited as measures of the effectiveness of e-learning system [29]. The researcher used TAM model proposed by [30] and Updated Information Systems Success Model developed by [31] to investigate the factors that determine the effectiveness of e-learning and propose a conceptual modal for this research. Besides, the researchers used self efficacy as one of the factor which could determine the effectiveness of e-learning because self-efficacy is a person's self-evaluation of his or her own capabilities regarding a specific course of action and it can predict the learners involvement in using and getting satisfaction of using the learning system [16].

A. Technology Acceptance Model (TAM)

This paper proposes and tests a conceptual model of e-learning Technology Acceptance Model (TAM) which was developed by [30] and extensively used and discussed by researchers in the previous literature in educational context. TAM model studies the relationship between the variables such as perceived ease of use and perceived usefulness, attitude and intention to adopt.

TAM is an information systems theory that models how users accept and use technology when they indulge in learning activities. According to [32] developed Theory of Reason Action (TRA) in psychology research and proposes a behavioral model where two beliefs such as perceived ease of use and perceived usefulness become primary predictors of use intension of a system. The TRA posits that individual behavior is driven by behavioral intention where behavioral intention is a function of an individual's attitude toward the behavior and subjective norms surrounding the performance of the behavior. In other words, it states that one's behavior and the intent to behave is a function of one's attitude toward the behavior and their perceptions about the behavior. Therefore, behavior is the function of both attitudes and beliefs. Meanwhile, TAM proposes that perceived ease of use and perceived usefulness of technology are predictors of user attitude towards using the technology, subsequent behavioral intentions and actual usage. Perceived ease of use was also considered to influence perceived usefulness of technology.

Three tier Technology Acceptance Model which incorporate multidisciplinary prospective such as Technology Acceptance Model (TAM), Social Cognitive Theory (SCT), theory of planned behavior (TPB) and motivation could influence the e-learners behavioral intention to use technology will make the e-learning effective [23]. He also described, individual attitude (individual characteristics) towards information technology and system quality influences users' affective, cognitive and behavioral intentions to use the technology. Although the model is mainly used to explain the adoption of technology within organization, the constructs of the models are meant to be general and can be used to educational context using e-learning. In addition, [32] described the variables used in TAM model as universal to any type of computer system and users. The overall conceptual model is illustrated in Figure 1.

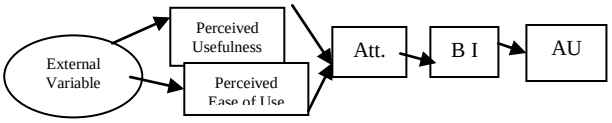


Figure 1: Technology Acceptance Model by Davis (1989)

However, it is unclear whether this model is still applicable for measuring its reliability, validity and generalisability in Malaysian context. Besides, TAM also suffers from other limitations such as inconsistencies in previous studies on the acceptance of technology [33]; [34].

Perceived Usefulness

According to [30] defined Perceived Usefulness as the degree to which a person believes using a particular system would enhance his or her job performance. Perceived Usefulness is a perception of students on the degree of improvement that they can gain from using e-learning system because of adopting the learning system.

Researchers used the TAM to measure the students' acceptance of technology in e-learning educational context. Most of the studies were about students' perceptions on using technologies which support Davis's TAM model [35] and [36]. Some of the researchers found that the variable

Perceived Usefulness was found to be the most influential variable in predicting the intention to use the e- learning technology [37]; [38]. The other research found that it has the influence on the intention to use the technology but it is not the most influential factor, besides there are other factors which plays important role [39]. Their research revealed that Perceived Usefulness became the determinant for the system usage. Besides, it also significant effect on the e-learners and improves their satisfaction of using the e-learning system [40].

Perceived Ease of Use

Perceived ease of use is defined as individual assessment of the extent interaction with specific information system or technology is free of mental effort or how effortless he or she perceives using the technology will be [30]. In this contact, many researchers believe that perceived ease of use determines the users' intention to adopt the new technology and use the system well [41]. This construct becomes one of the predominant determinant in TAM model as well as revised TAM models.

Perceived ease of use is students' perception of the ease of adopting an e-learning system. If the students perceive that e-learning technology is easy to use, then they will have positive attitude to use the system well and effectively. If the learners perceive an e-learning technology to be easy to use, they would also perceive it to be useful [42].

Updated Information Systems Success Model (Users' Satisfaction)

This was further developed by [31]. Obviously, this model has been used by many researchers to explain the antecedent factors which influence the users' satisfaction level when they use an electronic system to purchase goods or services. The model studies the relationship between information quality, system quality and service quality to predict the users' satisfaction and intention to use/use of the system with respect to the benefit the users gain after using the system.

Following the publication of DeLone and McLene's work, the original Information Success (IS) model received several criticisms from many scholars. Due to the criticism they reviewed and revised the model and produced an updated model in 2003. The original model consists of information quality, system quality, use, user satisfaction, individual benefit and organizational benefit. Later, they replaced individual benefit and organizational benefit to net benefit, added a new construct so called service quality and added intention to use to use construct. The ISS model is shown in figure 2 below.

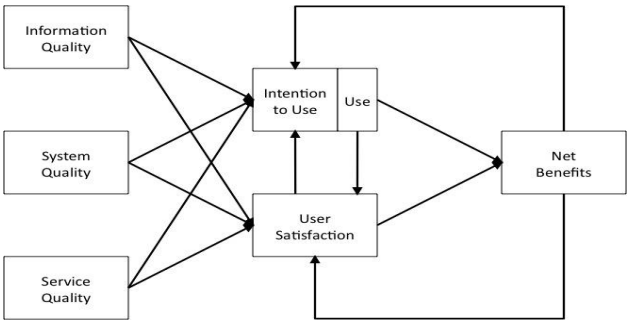


Figure 2: Information Systems Success Model by Delone and McLean (2003)

The researcher tend to use this model as a theoretical model in educational context to explain the degree of satisfaction the e-learning students achieve using e-learning system to acquire knowledge. The researcher replaced the construct net benefit with effectiveness of e-learning since the study is concerned about the factors that determine the effectiveness of e-learning. In addition, the construct service quality was omitted from the research model because many researchers found that it is much related to system quality. Besides, the researcher omitted the intention to use and only using use construct as the researcher's model because e-learning is mandatory in Malaysian government schools.

Users' satisfaction

Users' satisfaction can be defined as an overall users' attitude towards a service provider [43]. In the field of human-computer interaction, user satisfaction is visualized as the expression of affection obtained from an interaction and are significant in students early engagement opportunities and experiences, especially in e-learning context. Learning achievements is important to learners because satisfaction in e-learning can influence the decision to continue or terminated the course [40]; [44]. Since previous research has suggested strongly that e-learners' satisfaction has a positive impact on future intentions to continue using a system [45].

Content Quality

Information quality is the measures of information system output [46]. Information is important for internet interaction among users as internet was primarily developed to exchange information. The research use content quality instead of using information quality because content would be more appropriate to use in the context of education. Content is one of the main characteristics of website design where it is significant in aspects of e-learning, since the content's selection, analysis, designs, and presentation produces a comprehensive and attractive learning experience. [47] Opined that learners prefer to read the content if the quality of the content is well suited the students' need and very presentable manner. It must be

well-organized, effectively presented, interactive, clearly written, in the right length, useful, flexible, and provide appropriate degree of breath. Besides, timely, relevant, and game based motivating content is instrumental in an e-learning program.

System Quality

In [48] concluded that information and system quality impact on satisfaction of the system. The System can be defined as a set of elements standing in interrelations. System plays an important role in e-learning. And, hence infrastructure of technology and technical support of e-learning system are contributing factors to the acceptance of e-learning and in order to create an excellent and acceptability of e-learning, it is important to state that the technology and the e-learning system must be well-maintained and always up-to-date from time to time [49]; [50].

Delivery method

Delivery method of e-learning is the medium to transmit the course content from the instructors to the students. The Delivery method of e-learning can be synchronous or asynchronous. [50]. Synchronous delivery method refers to real-time and instructor-led e-learning, where all learners receive information simultaneously and communicate directly with other learners. Examples include teleconferencing (audio, video, or both), Internet chat forums, and instant messaging. Whereas, the asynchronous delivery method refers to transmission and receipt of information of contents which does not occur simultaneously. The learners have to self-pace their own learning without instructors online. They communicate through e-mail, online bulletin boards, newsgroups, and Weblogs. It is feedback technology. The students in this research use the asynchronous method to learn. .

Self-Efficacy

Self-efficacy, defined as a "belief in one's capabilities to organize and execute the courses of action required to manage prospective situations" which affects both cognitive and affective dimensions of learning processes. Students can construct their self-efficacy beliefs through four different sources of experiences: mastery experiences, vicarious experiences, verbal persuasion, and physiological and affective states [24]. It also can be defined as a person's self-evaluation accomplishing a task using a computer. E-learners with high self-efficacy will be more satisfied using e-learning system compared with those with relatively low self-efficacy [44]. High self-efficacy helps students become persistent in pursuing intrinsic goals and willing to attempt difficult tasks.

B. The proposed research model

The objective of this study is to investigate the factors that determine the effectiveness of e-learning among the secondary school e-learners in Malaysia. In this model, the researcher used two constructs, namely Perceived Ease of Use and Perceived Usefulness from TAM model [30]. This is to observe how much these two constructs influence the use of e-learning system and users' satisfactions and subsequently influence the effectiveness of e-learning among the e-learners. Whereas, the other constructs such as Content Quality, System Quality, Use were adopted form ISS model and the benefit construct form ISS model was changed to Effectiveness of e-learning since the research is conducted in education. Lastly, Delivery Method and Self Efficacy were added to this model to observe whether they have any influence on the effectiveness of e-learning.

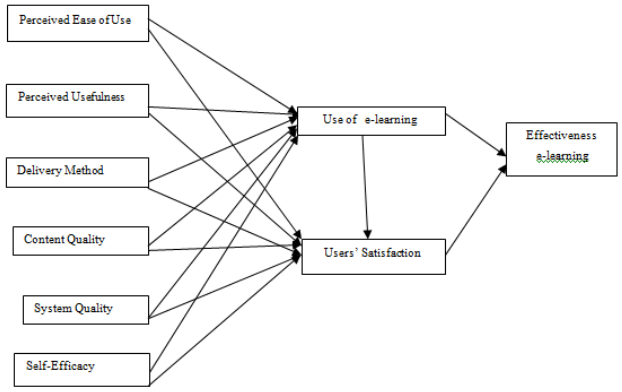


Figure 3: Proposed research model

The proposed conceptual model is developed based on the theoretical framework of previous research models namely, TAM model proposed by [30] and Updated Information Systems Success Model developed by [31]. It is proposed to measure the effectiveness of e-learning among students in Malaysian secondary schools. For the purpose of this study, the two variables which were used in TAM namely Perceived Ease and Perceived Usefulness are used to study their influence on the use e-learning system and their satisfaction level which in turn, influence the effectiveness of e-learning among students in secondary schools in Malaysia. The researcher focused on the perceived ease of use and perceived usefulness variable as these variable shown in TAM model shown significant influence on the learners' satisfaction [51].

In addition to this, Self Efficacy and Delivery Method are added to the model to test their influence on the use of e-learning and users' satisfaction and subsequently on the effectiveness of e-learning. The proposed research model is depicted in Figure 3.

III. CONCLUSION

This proposed model is foresee to solve the e-learning problem faced by students. This include difficulties in using the e-learning system to find the relevant information that affects their satisfactions and deprive them from using the system and in turn affects the e-learning effectiveness. Difficulty to understand the contents often force them to refer to their teachers or peers to get explanation about the contents. Due to this, they get bored with the system easily. Consequently, they opt to look for information manually from the books or refer to their teachers or peers. This implicates their perception about the quality of the contents, usefulness of e-learning system, ease of finding needed information, satisfaction and using the e-learning system. The students who use the e-learning system in the school often found that they couldn't have access to large files such as audio and video files; could not get the needed information in time because the delivery of information is taking quite long, subsequently, they feel disappointed with the system and decide not to use the e-learning system. They prefer to engage themselves with traditional learning mode. Failure in using the e-learning system will resulted in depriving their benefit of using the system to gain knowledge and improvement in their educational endeavours.

The learning contents which are prepared by the teachers are delivered through asynchronous method which means the transmission and receipt of the learning contents do not occur simultaneously. VLE FROG provides the means for the delivery of learning contents through web pages whereby the students need access to the web site and locate the web page where they can read and do their assignments given by their teachers. They often face difficulties accessing the sites using their ID and this unable them to indulge in e-learning, therefore, affects their learning activities and their performance in their studies. Strong accessibility of delivery method system could give best and engage students, resulting to educational success. Self-efficacy if found to be less makes students feel they are less confidence whenever they use e-learning system to learn, having preference to their teachers' presence whenever they indulge in e-learning process and get more comfortable with the traditional face-to-face teaching and learning method. Students with less self-efficacy tend to degrade the students achievement in their studies. And those with good self-efficacy prefer using internet to learn and having confidence in using the e-learning system. In general, these problems or factors affect the use of e-learning system and users' satisfaction of using the system which subsequently affect the students' effective learning.

References

[1] A. Paradise, State of the Industry Report, Rep. Alexandria: ASTD Research Department, 2008.

[2] A. Sangrà and D. Vlachopoulos, "E-Learning definition," UOC, 2010.

[3] H.M. Selim, "Critical success factors for e-learning acceptance: Confirmatory factor models," Computers & Education, vol. 49, pp. 396-413, 2007.

[4] T. Urdan, C. Weggen, and W. Hambrecht, "Corporate e-learning: Exploring a new frontier," WR Hambrecht, 2000.

[5] B.H. Khan, "Managing e-learning: Design, delivery, implementation, and evaluation," Hershey PA', Information Science Publishing, 2005.

[6] M.D. Roblyer, L. Davis, S.C. Mills, J. Marshall, and L. Pepe, "Toward practical procedures for predicting and promoting success in virtual school students," The American Journal of Distance Education, vol. 22, pp. 90–109, 2008.

[7] C. Weiner, "Key ingredients to online learning: Adolescent students study in cyberspace – the nature of the study," International Journal on E-Learning, vol. 2, pp. 44–50, 2003.

[8] R. Dewstove, and N. Wright, "Secondary school students, online learning, and external support in New Zealand," Computers in the Schools, vol. 22, pp. 111–122, 2005.

[9] S. Tunison, and B. Noonan, "On-line learning: Secondary students' first experience," Canadian Journal of Education, vol. 26, pp. 495–511, 2001.

[10] D. Asirvatham, "E-learning initiatives in Malaysian E-learning: The future" http://www.marketresearch.com. 25th January 2005.

[11] R. Abdul Rahman "E-learning initiatives in Malaysian schools," Presented at the third seminar workshop of the 7th programming cycle of APEID Activities, September 2004.

[12] P. Lowenthal, and B.G. Wilson, "Labels do matter! A critique of AECT's redefinition," 2010.

13] T. Volery, and D. Lord, "Critical success factors in online education," The International Journal of Education Management, vol. 14, pp. 216 – 223, 2000.

[14] J. Rossiter, "An e-learning vision: Towards a Pan-Canadian strategy and action plan," (CANARIE Discussion paper). Ottawa, ON: CANARIE Inc. Retrieved August 15, 2005.

[15] R. Kurtus, "What is E-learning?" http://www.school-for-champions.com, 2004.

[16] G. Piccoli, R. Ahmad, and B. Ives, "Web-based virtual learning environments: a research framework and a preliminary assessment of effectiveness in basic IT skill training," MIS Quarterly, vol. 25, pp. 401–426, 2001.

[17] D. Ely, "Condition that facilitate the implementation of education technology innovation" Education technology, vol. 39, pp. 23-27, 1999.

[18] L. Siragusa, "Research into the effectiveness of online learning: Survey finding," Proceeding Western Australian Institute for Education Research Forum, April 2002.

[19] C.A. Philip, "A Review of e-Learning in Canada: A Rough Sketch of the Evidence, Gaps and Promising Directions," Canadian Journal of Learning and Technology vol. 32, pp. 3-10, 2006.

[20] B.J. Zimmerman, and D.H. Schunk, "Self-regulated learning and academic achievement: Theory, research, and practice," New York, NY: Springer-Verlag, 1989.

[21] K. Swan, "Virtual interaction: Design factors affecting student satisfaction and perceived learning in asynchronous online courses," Distance Education, vol. 22, pp. 306–331, 2001.

[22] K Passerini, and M. Granger, "A development model for distant learning using the Internet," Computer and education, vol. 34, pp. 1-15, 2000.

[23] S. Liaw, "Understanding computers and the internet as a work assisted tool," Computer in human behavior, vol. 23, pp. 399-414, 2007.

[24] A. Bandura, Self Efficacy: The exercise of Control, New York: W.H.Freeman and company, 1997.

[25] K. Amoako-Gyampah, "Perceived usefulness, user involvement and behavioral intention: an empirical study of ERP implementation," Computers in Human Behavior, vol. 23, pp. 1232-1248, 2007.

[26] S.B. Eom, H.J. Wen, and N. Ashill, "The determinants of students' perceived learning outcomes and satisfaction in university online education: An empirical investigation. Decision Sciences," Journal of Innovative Education, vol. 4 pp. 215–235, 2006.

[27] J.W. Peltier, W. Drago, and J.A. Schibrowsky, "Virtual communities and the assessment of online marketing education," Journal of Marketing Education, vol. 25, pp. 260–276, 2003.

[28] W. Chih-Hsuan, M.S. David, and E.R. Margaret, "Do they really dislike social studies? A study of middle school and high school students," Journal of Social Studies Research, vol. 28, pp. 16–26, 2013.

[29] M. Alavi, B.C. Wheeler and J.S. Valacich, "Using IT to reengineer business education: An exploratory investigation of collaborative telelearning," MIS Quarterly, vol. 19, pp. 293–312, 1995.

[30] F.D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," MIS Quarterly, vol.13, pp. 319-340, 1989.

[31] W.H. DeLone, & E.R. McLean, "The Delone and McLean Model of Information Systems Success: a ten-year update," Journal of Management Information Systems, vol. 19, pp. 9-30, 2003.

[32] C. Yi-Cheng, "Predicting College Students' Use of E-Learning Systems: an Attempt to Extend Technology Acceptance Model," 2007.

[33] S. McCoy, D.F. Galletta, and W.R. King, "Integrating national culture into IS research: the need for current individual level measures," Communications of the Association for Information Systems, vol. 15, pp. 12, 2005.

[34] V. Venkatesh, M.G. Morris, G.B. Davis, and F.D. Davis, "User Acceptance of information technology: towards a unified view," MIS Quarterly, vol. 27, pp. 425-478, 2003.

[35] A. Rai, S. Lang, and R. Welker, "Assessing the Validity of IS Model: An empirical test on theoretical analysis," Information System Research vol. 11, pp. 50-69, 2002.

[36] M.A. Mahmud, J.M. Burn, L.A. Gemoets, and C. Jacquez, "Variables affecting information technology end-user satisfaction: a meta-analysis of the empirical literature," International Journal of Human-Computer Studies, vol. 52, pp. 751-771, 2000.

[37] S.C. Chang, and F.C. Tung, "An empirical investigation of students' behavioural intentions to use the online learning course websites," British Journal of Educational Technology, vol. 39, pp. 71-83, 2008.

[38] R.G. Saadé, and I. Galloway, "Understanding intention to use multimedia information systems for learning," International Journal of an Emerging Transdiscipline, vol. 2, pp. 287-296, 2005.

[39] T. Chesney, "An acceptance model for useful and fun information systems," 2006.

[40] C.M. Chiu, M.H. Hsu, S.Y. Sun, , T.C. Lin, and P.C. Sun, "Usability, quality, value and e-learning continuance decisions" Computers & Education, vol.45, pp. 399-416, 2005.

[41] R. Agarwal and E. Karahanna, "Time flies when you're having fun: cognitive absorption and beliefs about information technology usage," MIS quarterly, vol. 24, pp. 665-694, 2000.

[42] B.J.L. Landry, R. Griffeth, and S. Hartman, "Measuring student perceptions of blackboard using the technology acceptance model," Decision Sciences Journal of Innovative Education, vol. 4, pp. 87-99, 2006.

[43] T Levesque, and G.H.G. McDougall, "Determinants of customer satisfaction in retail banking," International Journal of Bank Marketing, vol. 14, pp. 12-20, 1996.

[44] Y. Levy, "Comparing dropouts and persistence in e-learning courses," Computers & Education, vol. 48, pp. 185–204, 2007.

[45] R. Oliver, and J. Herrington, "Exploring technology-mediated learning from a pedagogical perspective," Interactive Learning Environments, vol. 11, pp. 111–126, 2003.

[46] W.H. DeLone, and E.R. McLean, "Information systems success: the quest for the dependent variable," Information systems research, vol. 3, pp. 60-95, 1992.

[47] D.Y. Shee, & Y.S. Wang, "Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications," Computers & Education, vol. 50, pp. 894–905, 2008.

[48] S. Negash, T. Ryan, and M. Igbaria, "Quality and effectiveness in Web-based customer support systems," Information & Management, vol. 40, pp. 757-768, 2003.

[49] L. Poon, and Yong The process of learning 3rd ed, Australia: Prentice Hall, 2004.

[50] D. Compeau, & C. Higgins, "Computer Self-Efficacy: Development of a Measure and Initial Test," MIS Quarterly, vol. 19, pp. 189-211, 1995.

[51] M. H. Fagan, S. Neill, and B.R. Wooldridge, "An Empirical Investigation into the Relationship between Computer Self-Efficacy, Anxiety, Experience, Support and Usage," Journal of Computer Information Systems, vol. 44, pp. 95-106, 2004.

Undergraduate Students E-Attendance System Using Fingerprint and SMS Technologies

An Appraisal to Improve Students and Parents Psychological Stability

Simeon A. Adeyemi
Dept. of Computer Science
Federal Polytechnic, Offa
Offa, Nigeria

Modinat A. Mabayoje, Abdullateef O. Balogun and Hammed A. Mojeed
Dept. of Computer Science
University of Ilorin
Ilorin, Nigeria
mabayoje.ma@unilorin.edu.ng,

***Abstract*—Attendance is the measure of presence or absence of a person or an individual in an organized formal group (institutions, organizations, meetings, and so on). The need for a time saving and efficient attendance process, in this era of time management and consciousness, which the present manual method of paper sheet attendance cannot satisfy, is the motivation for this research work. With attendance taking a crucial part of any serious academic requirements, this research work integrates the use of fingerprint biometrics to automate all students' attendance processes (exams, lectures and entrance) by combining minutiae extraction algorithm based on Midpoint Ridge Contour Representation (MRCR), for extraction of minutiae information, with Fingerprint Recognition Minutiae Score Matching (FRMSM) Algorithm which is based on Minutiae Matching Score to match the test fingerprint to the template stored in the database. Also, included is an SMS feature which sends a message to parents upon student's entry and exit into school premises.**

The fingerprint image is captured using a reader and the minutiae features are extracted and stored in database against which another features are compared for authentication. The system is implemented using Visual Basic.Net programming language.

Keywords—E-attendance; Bio-metrics; Fingerprint; Minutiae; SMS, Authenticayion.

I. INTRODUCTION

Organizations of all sizes use time and attendance systems to record when employees start and stop work and when the work is performed. In addition to keeping tracks of when employees work, it also helps to keep tabs on when employees are not working. Education is another crucial area where attendance of students and staff is given serious emphasis, such that it forms part of the requirements and metrics for determining students' performance within the academic activities. Attendance therefore, becomes an inseparable component of any serious academic environment, hence, the need for this research. Electronic Attendance (E-Attendance) is the process of taking the record of presence or absence of an individual within an organized formal group with the use of computer gadgets. The adoption of E-Attendance as opposed to other methods of taking attendance is preferred on the basis of its advantages premised on computer processing speed, accuracy, throughput, storage capacity, efficiency and flexibility.

Various tools available for implementing this include electronic tags, barcode badges, magnetic stripe cards, biometrics (vein reader, hand geometry, fingerprint, or facial), and touch screens.

Bio-metrics is the measurement and statistical analysis of people's physical and behavioral characteristics with a view to identifying and authenticating such an individual (Mike, 2013). Fingerprint Biometric is used to implement this attendance system, due to its universality and unique identity, combined with the use of SMS as a mode of communicating student's attendance at school to parents/guardians. This will improve students' performance and boosts parents' peace of mind on their wards' where about. Improved performance will eliminate previously failing students' psychological instability. And parents are more fully aware not only of their wards' presence but also absence from school.

As paramount as attendance is to academic performance of students, it is ironic that institutions face, on daily basis, a dwindling number of attendees in lectures. The current adopted manual method of taking attendance on paper has a lot of peculiar problems that makes it inefficient as the appropriate means, such problems include, Wastage of quality lecture time, Room for impersonation, Loss of sheet in transit, Cumbersome and stressful to monitor especially when students' population is large, and so on. More so, due to low attendance, performance deteriorates and as a result, more students suffer psychological instability resulting from these failures. It is convincing therefore that the present method cannot solve the existing associated problems, hence, the need for a transition from manual to electronic. This research is conducted to solve these problems using Electronic Attendance System, by integrating the use of Fingerprint biometrics and SMS technologies.

The main aim of this research is to propose an Electronic Attendance System that combines fingerprint Biometrics to uniquely identify students for attendance and SMS Technologies to solve all students' attendance-related problems, during lectures, exams and entry and exit of students into school and to ensure parents/guardians awareness of their wards' seriousness and presence at school via SMS communication as an attempt to improve students and parents psychological stability arising from the system's performance.

II. METHODOLOGY

Data for this project were obtained from direct interview of students and staff of Federal Polytechnic, Offa, transcription from records, and the Internet. For the implementation of the project, Microsoft Visual Basic.NET 2013 version, MySQL (Version 5.6.12.2) for the database and Fingerprint SDK 2009 by Grianule as well as AT Commands for communicating with SMS-supported devices were adopted.

A total of 200 students and 50 staff of the Federal Polytechnic, Offa, Nigeria, were interviewed. The table below shows the summary of the questions asked during the interview as regards the effect of E-Attendance under the following key points. Each number represents the percentage of the total number of respondents interviewed.

TABLE I. SUMMARY OF THE DIRECT INTERVIEW CONDUCTED FOR STUDENTS AND STAFF OF FEDERAL POLYTECHNIC, OFFA ON THE EFFECT OF E-ATTENDANCE ON PERFORMANCE.

Questions	**Students**				**Lecturers**			
	S	***A***	***D***	***SD***	***S***	***A***	***D***	***SD***
Positive effect on Performance	86 %	9%	5%	0%	90%	7%	3 %	0%
Motivation of student's discipline and punctuality	76 %	15 %	8%	1%	82%	16%	2 %	0%
Students at risk of poor academic performance will continue to successfully persist	60 %	25 %	12 %	3%	75%	20%	5 %	0%
Due to availability, thumbprint will be better and simpler to use for taking attendance.	45 %	35 %	17 %	3%	53%	32%	14 %	1%
E-Attendance will save quality lecture time	60 %	20 %	18 %	2%	75%	20%	5 %	0%
E-Attendance record will become more easily accessible					73 %	21 %	6 %	0 %
E-attendance will eliminate issue of record loss					65 %	23 %	12 %	0 %

[b.] S – Strongly Agree A – Agree D – Disagree SD – Strongly Disagree

A. The Research Design

This project combined minutiae extraction algorithm based on Midpoint Ridge Contour Representation (MRCR) [2], as in Fig. 1, for extraction of minutiae information, with Fingerprint Recognition Minutiae Score Matching (FRMSM) Algorithm which is based on Minutiae Matching Score to match the test fingerprint with the template stored in the database [26], as in Fig. 2. Shown the figures below are the two algorithms.

Step1: Segmentation: To separate foreground from background of fingerprint image. A 64 x 64 region is extracted from fingerprint image.
Step2: Normalization: The grayscale intensities in 64 x 64 regions are normalized to a constant mean and variance to remove the effects of sensor noise and grayscale variations due to finger pressure differences.
Step3: Filtering (Gabor Filter): After the normalization, the contrast of the ridges is enhanced by filtering 64 x 64 normalized windows by appropriately tuned Gabor filter.
Step4: Scanning: Processed fingerprint image is then scanned from top to bottom and left to right and transitions from white (background) to black (foreground) are detected.
Step 5: Calculate Vector: The length vector is calculated in all the eight directions of contour. Each contour element represents a pixel on the contour, contains fields for the x, y coordinates of the pixel.

Fig. 1. Bupesh MRCR algorithm. [2].

The method takes less and do not detect any false minutiae [2].

The FRMSM Algorithm on the other hand is as in Fig. 2.

FRMSM Algorithm By (Ravi, Raja and Venugopal, 2009)
Input: Gray-scale Fingerprint image.
Output: Verified fingerprint image with matching score.
Step 1: Binarize Fingerprint image
Step 2: Thin the binarized image
Step 3: Extract Minutiae points.
Step 4: Generate Data matrix to get the position, orientation and type of minutiae.
Step 5: Match test fingerprint with template in database.
Step 6: Compute Matching Score of the two images.
Step 7: if matching score is 1, images are matched, else if 0, then they are mismatched.

Fig. 2. FRMSM algorithm [26].

B. *System Structural Descriptions*

The project is made up of five (5) modules namely:

1) Registration Module: This module handles all forms of registrations required from students and Lecturers. All students are required to register their details, including their thumbprint during registration, using an interface and a fingerprint reader device for the data collection.

2) Entrance Module: Students coming into or going out of the school premises are required to thumbprint for authentication. Once authenticated, student's detail is displayed, with a passport, and an SMS is sent across to students' parents/guardian's phone, stating the name of such student and the date and time of arrival to or departure from school premise. Student who cannot be authenticated is brought to the notice of security personnel for questioning.

3) Lecture Attendance Module: This handles Lecture attendance. A Lecturer supplies his/her staff ID, courses offered by such lecturer are loaded. A course is selected and all registered students for the course are automatically loaded. As each student thumbprints and is authenticated, his/her Matriculation number moves to Attendance List. Lecturer can submit attendance list to the database when he deems it fit during the lecture. Reports of such attendance over a given period or specified interval can be obtained and printed.

4) Examination Admittance and Attendance Module: This module handles similar task as Lecture attendance module but information required such as exam date, exam time, duration and venue are required from invigilators before attendance is taken. Each authenticated student will have his/her details displayed, including passport.

4) SMS Module: This Module is responsible for encapsulating all SMS features of the system. It involves everything that has to do with sending SMS to student's parent/guardian at entry into and exit from school premises, as well as sending of bulk SMS to all parents or all registered staff by the School Management.

C. *Conceptual Model of The System*

The system's conceptual model is shown in Fig. 3. Below.

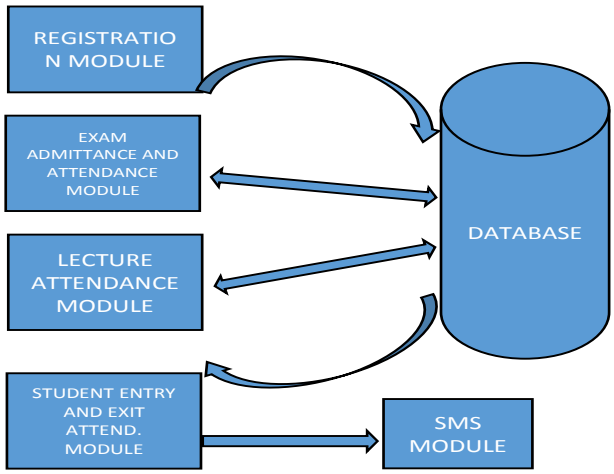


Fig. 3. System conceptual model

D. *Use Case Diagram for The System*

The use case diagram of the system is as shown below:

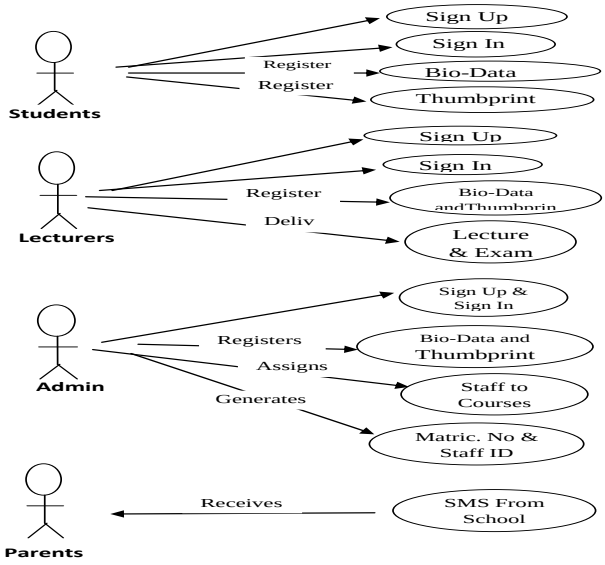


Fig. 4. Use-case diagram of the system.

III. RESULT AND DISCUSSION

After the system was developed, it was run and tested on Windows 7 Operating System environment, installed on Intel(R) Core(TM) i5 – 2540M CPU with 4GB RAM size and operating at a frequency of 2.60GHz. MySQL (Version 5.6.12.2) database was used to store various data obtained from students and staff as well as the minutiae information extracted from the fingerprint data. And it was tested using DigitalPersonaU.Are.U 4000B Fingerprint reader. Some of the output interfaces obtained are as shown in Fig. 5 – Fig. 9 below.

Fig. 5. Main page

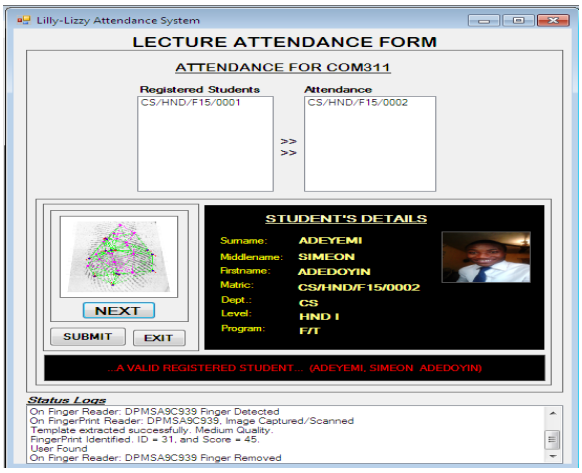


Fig. 6. Form showing lecture attendance in progress I

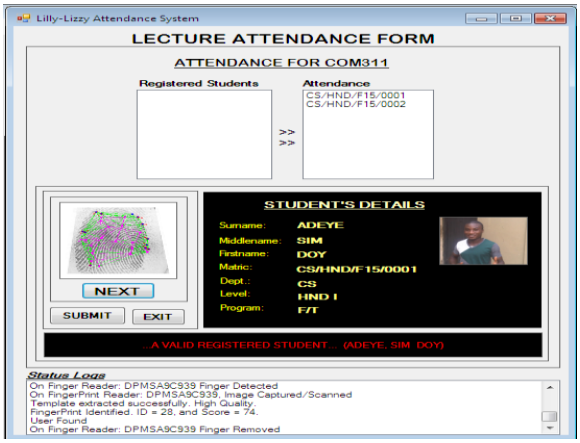


Fig. 7. Form showing lecture attendance in progress II

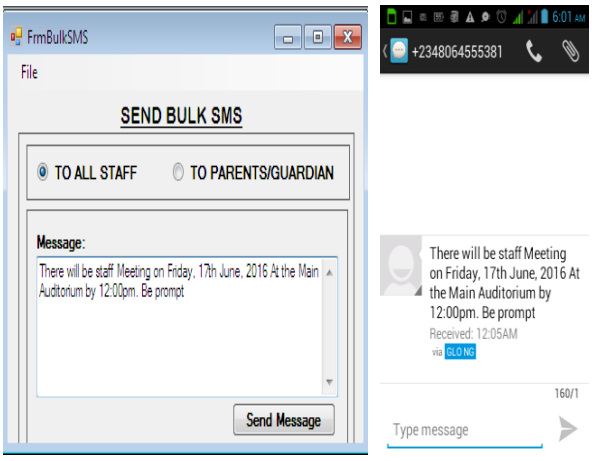


Fig. 8. Sample bulk SMS sent from school to all staff

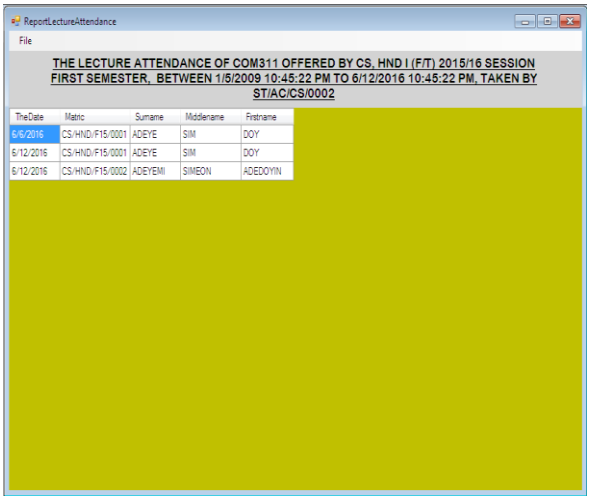


Fig. 9. Lecture attendance report over a specified time interval.

IV. CONCLUSION

Over the years, Attendance in institutions had been done manually. And existing electronic attendance systems only handle a particular attendance requirement and without a feedback to the parents, bye and large, for psychological purpose. This research however incorporates a holistic approach to solving all attendance-related issues for undergraduate students and helps to manage effective involvement of parents in collaborative effort to monitoring students' level of participation in school. From the results obtained, it is evident that the proposed system will facilitate academic excellence and improve standard of education as lecture delivery per student will, by implication, increase. It will also reduce health issues associated with fear of failure by students, as students gain more self-confidence with more lecture attendance, as such, they stand a better chance of performing better. Parents also enjoys peace of mind knowing their wards attend school regularly and with knowledge of the wards' whereabouts, hypertensive parents will live better, with such peace. The system will definitely improve Psychological stability of students and their parents alike. It encourages a collaboration of institutions and parents in helping to monitor and instigates students' seriousness academically.

Although, the use of fingerprint Biometrics technology for attendance is not too new, there has not been any of its application integrated with Short Message Service (SMS) technology in Nigerian institutions. Also, at the moment no higher institution has applied this technology ever for students' attendance. Among the few higher institutions using fingerprint biometrics attendance system, it has only been used in taking staff attendance, with focus on ensuring quantifiable service of staff to determine payment. This research however, focuses on students' attendance with a view to solving examination malpractice through

impersonation, lecture attendance associated problems (for increased lecture throughput, by saving the time lecturers spend monitoring students during attendance, also for the authenticity of such attendance) and instigate seriousness of students by ensuring parental/guardian monitoring of student's level of participation or presence and absence in school.

REFERENCES

[1] K.J. Anil, R. Arun, and P. Salil, "An introduction to Biometric recognition," IEEE Trans. on Circuits and Syst. for Video Tech., vol. 14, Issue 1, January 2004, pp. 4 – 20

[2] G. Bhupesh, , T.K. Bandopadhyaya, and S. Sudhir, "Fingerprint feature extraction using Midpoint Ridge Contour Method and Neural Network," Int. J. of Comp. Sci. and Netwk Sec. IJCSNS, vol. 8, no. 7, pp. 99-109, 2008.

[3] E. Burd, and B. Hodgson, "Attendance and attainment: a five year study," Innov. in Teaching and Learning in Info. and Comp. Sci., vol. 5 no. 2. Retrieved from: http://journals.heacademy.ac.uk/doi/abs/10.11120/ital.2006.05020004. Accessed: January 16, 2016.

[4] L. Chien, "A survey of Biometrics security systems," http://www.cse.wustl.edu/~jain/cse571-11/ftp/biomet.pdf. Accessed: January 25, 2016, [n.d].

[5] J. Colby, "Attendance and attainment – A comparative study," Presented at the 5th Ann. Conf. of ICS-LTSN, Univ. of Ulster, 31st August to 2nd September, 2004.

[6] A. Davis, "The correlation between attendance and achievement," Teaching Fellowships, paper 13, Dublin Institute of Tech., 2011. http://arrow.dit.ie/fellow/13. Accessed: January 25, 2016.

[7] B. Debbie, B. Chris, and C. Jane, "Intervening early: Attendance and performance monitoring as a trigger for first year support in the Biosciences," Bioscience Education, June 16, 2010. Retrieved from: http://journals.heacademy.ac.uk/doi/abs/10.3108/beej.15.4. Accessed: January 16, 2016.

[8] B. Debra, C. Daniel, B. Stacy, and R. Collin, "All or nothing? Midpoint credentials for students who stop short of the Baccalaureate degree," Univ. of Illinois at Urbana-Champaign, November 15, 2011, pp. 9-10.

[9] Developer's home, "SMS tutorial: Introduction to AT commands, basic commands and extended commands." 2016. http://www.developershome.com/sms/atCommandsIntro.asp. Accessed: November 14, 2015.

[10] Digital Persona, Inc., "Guide to fingerprint recognition," 720 Bay Road Redwood City, CA 94063 USA, [n.d]. http://www.digitalpersona.com. Accessed: September 15, 2015.

[11] European Telecommunications standards Institute [ETSI], "GSM document 02/82." http://www.etsi.org/standards. Accessed: June 6, 2016.

[12] B.M. Farahanum, "Student's attendance system using fingerprint technology," B.Sc. Project, Data Comm. And Networking Dept, Fac. of Info. Tech. And Quantitative Sci., Mara Univ. of Tech., Shah Alam, 2006.

[13] G.A. Graevenitz, "Introduction to fingerprint technology," 2003. http://info.bioenabletech.com/training/fingerprint/introduction%20to%20fingerprint.pdf. Accessed: September 15, 2015.

[14] R. Heidn, "A world history of fingerprint," Chinese People Public Sec. Univ. Press, ISBN: 978-7-81109-789-4, 2008.

[15] R. Josphineleela, and Dr. M. Ramakrishman, "An efficient automatic attendance system using fingerprints reconstruction techniques," Int. J. of Comp. Sci. and Info. Sec., IJCSIS, vol. 10, no. 3, 2012.

[16] S. Kadry, and M. Smaili, "Wireless attendance management system based on iris recognition," Scientific Research and Essays, vol. 5, issue 12, June, 2010, pp. 1428-1435.

[17] G. Kanimozhi, and T. Chakravarthy, "Secure biometric authentication architecture using fingerprint using FAR," in Int. J. of Engr. Research and Tech., IJERT, vol. 2, issue 2, February 2013, ISSN: 2278-0181.

[18] S. Khalid, and M. Pejas-Romuald, "Biometrics, Computer security, Systems and Artificial Intelligent applications", Berlin: Springer-Verlag, 2006, ISBN 0387362320.

[19] P.G. Lakshmi, M. Pandimadevi, P.G. Ramu, and P. Ramya, "Implementation of attendance management system using SMART-FR," in Int. J. of Adv. Research in Comp. and Comm. Engr, vol. 3, issue 11, November, 2014.

[20] P. Loucopoulos, and R. Zicari, "Conceptual modeling, databases, and CASE: An integrated view of information systems development," New York: John Wiley & Sons, Inc., 1992.

[21] W. Marguerite, "Libya's election ushers in new voter tech - World Policy Institute," in World Policy Blog, 2014. http://www.worldpolicy.org/blog/2014/06/25/libyas-election-ushers-new-voter-tech. Accessed: June 6, 2016.

[22] M.A. Mohammad, and A. Norhayati, "A short message service for campus wide information delivery," in 4th Nat. Conf. on Telecomm. Tech. proceedings, Shah Alam, Malaysia, 2003, pp216-221.

[23] S. Nurbek, and G. Selim, "Attendance control system based on RFID-technology," Int. J. of Comp. Sci., IJCSI, vol. 9, Issue 3, no 1, May, 2012, ISSN (Online): 1694-0814, pp. 227.

[24] O. Peter, O. Anne, W. Shaun, and K. Lucy, "Biometrics security system," 2011. Retrieved from: http://www.findbiometrics.com. Accessed: September 15, 2015.

[25] Portio Research, "Analysis and forecasts for Mobile Messaging worldwide - PortioResearch". http://www.portioresearch.com/en/messaging-reports/mobile-messaging-research/mobile-messaging-futures-2014-2018.aspx. Accessed: November 15, 2015.

[26] J. Ravi, K.B. Raja, and K.R. Venugopal, "Fingerprint recognition using minutia score matching," in Int. J. of Engr. Sci. and Tech., IJEST, vol.1, no. 2, pp. 35-42, 2009.

[27] J. Rizzitano, "Understanding Mobile Modem SMS," 2006. http://www.forums.whirlpool.net.all>forumarchive>networking. Accessed: September 15, 2015.

[28] V. Shehu, and A. Dika, "Using real time computer vision algorithms in automatic attendance management systems", Proc. of the 32nd Int. Conf. on Info. Tech. Interfaces, ICITI, June, 2010, Cavtat, Croatia.

[29] K. Silver, "OMG: Text messaging turns 19 this week ... and this is the Brit we have to thank for our sore thumbs". Daily Mail (London), December 7, 2011. http://www.dailymail.co.uk/news/article-2070892/OMG-Text-messaging-turns-19-Meet-Neil-Papworth-Brit-thank.html. Accessed: June 6, 2016.

[30] T.A. Tomi, and M. Alan, "Time to confirm some mobile user numbers: sms, mms, mobile internet, M-News," January,2011. http://communities-dominate.blogs.com/brands/2011/01/time-to-confirm-some-mobile-user-numbers-sms-mms-mobile-internet-m-news.html. Accessed: June 3, 2016.

[31] K.K. Veena, and V.M. Thakare, "Short message service using sms gateway," in Int. J. on Comp. Sc. and Engr., IJCSE, vol. 2, no. 4, 2010, pp. 1487-1491.

[32] Wikipedia, "Short message service," 2016. https: //en.wikipedia.org/wiki/Short_Message_Service. Accessed: June 6, 2016.

Development of Second-Level Authentication Process for a Higher Level Security in ATM Transactions

J. K. Alhassan
Department of Computer Science
Federal University of Technology
Minna, Nigeria
E-mail: jkalhassan@futminna.edu.ng

A. Ochoche
Statistics Department,
Central Bank of Nigeria,
Headquarters, Abuja, Nigeria
oabraham@cbn.gov.ng

A. M. Enagi,
Department of Information and Media Technology,
Federal University of Technology,
Minna, Nigeria

G. O. Shefiu
Department of Information and Media Technology,
Federal University of Technology, Minna, Nigeria

B. L. Muhammad-Bello
Department of Information and Media Technology,
Federal University of Technology, Minna, Nigeria

***Abstract*— The increase of automated teller machine (ATM) frauds has actuated the development of new authentication mechanisms to overcome security problems of personal identification numbers (PIN). As a rule, ATMs give users three tries for authentication to the bank server. In an event where the user fails to authenticate to the bank server, the ATM card will typically be blocked and also confiscated by the ATM. In this research, a model to provide second-level authentication for ATM transactions was proposed and implemented using Visual Studio and MySQL. To operate this system, the user is to insert the ATM card into the machine, provide the PIN and then set the withdrawal limit. Any amount above this limit will request the user to enter the second level authentication code which is sent to the user's mobile phone by the bank server. The application was tested and found to solve the problem of cards confiscation and stressful existing process of retrieving such cards. It is recommended for implementation by banks and other financial institutions.**

Keywords- authentication; security; transaction; automated teller machine

I. INTRODUCTION

The earliest and old-fashioned society lacked to some extent monetary tools, therefore, the whole give-and-take of goods and commodities was accomplished by the "barter system" (1). The contemporary society though started using monetary tools as a unit of give-and-take which now substituted the barter system. Therefore, money in numerous values were now used as the solitary buying power as contrary to the barter system. The modern age has substituted these old-style financial tools from a paper and metal centered money to "plastic money" in the form of credit cards, debit cards, and so on (2). This has caused in the cumulative usage of Automated Teller Machine (ATM) all over the world. The statistics of ATM card owners has persistent to raise as consequence of e-payment consciousness and the placement of extra ATM cash points by banks all everywhere the world. In Nigeria, 80-90% of all e-banking dealings are conceded out via the ATM channel (1). Paradoxically, actions of card impostors and 'intelligent' criminals seem to be on the rise. ATM scam has developed a countrywide subject touching together the banks and customers (3). Numerous banks have sustained to caution ATM card operators against revealing their ATM card particulars to a second party in order to impose the safety of ATM usage. Common approaches used by cheats to propagate ATM scam comprise, absolute card robbery, shoulder surfing of operators at ATM points, usage of false personal identification number (PIN) Pad overlap and PIN interference via emails and text messages.

The difficult of ATM scam does not touch only the banks, somewhat, it is a large risk to all parties involved and it needs a synchronized and supportive action on the part of the bank, bank clients and the law implementation organizations (4), (5). ATM scams do not just cause economic harm to banks but they likewise demoralize clients' self-assurance in the usage of ATMs (6). This would dampen a larger usage of ATM for financial dealings. Further so, ATM facilities are extremely

This research was funded by Tertiary Education Trust Fund institution-based research intervention fund (TETFUND/FUTMINNA/2014/47) through Federal University of Technology, Minna, Nigeria.

lucrative for banks and numerous banks particularly in Nigeria violently market the usage of ATM cards (7). It is consequently in the importance of banks to avert ATM scams. Therefore, safety and insurance actions that provides better defense to the ATMs must to be put in place.

To guarantee the protection and honesty of internal online payments, many banks have presented a second level authentication method to confirm online dealings, (8) and (9). This technology was first used by Google to improve the safety of electronic mail account holders. Second level authentication also known as two factor authentication or two-step confirmation was used by Google to add an additional cover of safety to operators' Google Apps accounts by demanding them to enter a verification code in addition to their username and password, while signing in to their account (10), (11), (12) and (13). Second level authentication is a security procedure that can be consummate by applying either a mobile phone (SMS) or token device which offers a one-time password for transaction authentication.

This authentication procedure aids to guard operators' accounts from illegal admittance. Consequently, should an illegal operator manage to get a operator's password/ PIN, or even if a password is broken, predicted, or else taken, the invader will not be capable to validate to the system without admittance to the operator's confirmation encryptions, which only the operator can get via their own mobile phone or token device. Consequently, this study is on the usage of second level authentication and extra security features for ATM deal. The continuing parts of this study comprises literature review, methodology, results, discussion and findings.

II. Literature Review

Numerous earlier researches connected to the safety and operation of the ATM dedicated on improving ATM safety through biometric method. Though, very tiny or no effort has engrossed on the procedure of the ATM from the viewpoint of ATM cards being confined in the ATM. A biometric policy by finger print measure was planned by Das and Debbarma, (14), for improving ATM safety in Indian E-banking structure. In a latest study by Prithika and Rajalakshmi, (15) proposed using the Iris Recognition and Palm Vein (IRPV) recognition technology to avoid card replication and crimes via the ATM. The Advanced Encryption Standard (AES) algorithm was accepted in (16) to increase the safety level of ATM Banking Systems. Finger print recognition in digital image processing was implemented in (17) for a suggested new business model which would improve ATM safety. In (18) nevertheless, the standing safety of the ATM system was improved by joining the fingerprints of operators into the bank's database as a means for additional authentication procedure. ATM based fingerprint confirmation was established and simulated for ATM procedures in imperative to lessening cheats connected with the use of ATM. A latest study piloted by (19) improved the safety of the ATM by a mixture of fingerprint and GSM technology. The fingerprint recognition technology and PIN verification was implanted into a GSM modem which was linked to a microcontroller to produce 4 digits one-time passcode which is to be directed to a operator's mobile number every time the operator tries to register the finger print image on the banking system. A alike study in (20) also established a sample of an improved ATM verification system by Short Message Service (SMS) confirmation and also piloted a usability trying of the offered system. A system which combines facial recognition technology into the individuality confirmation procedure used in ATMs was offered in (20). The offered ATM model established using facial recognition software was more dependable in providing safety. A Novel Hybrid Technology in ATM Security Using Biometrics was suggested in (21). The suggested algorithm embraced in the study provided two stages of safety by together biometric and GSM technology. In adding to the PIN delivered, a second level authentication by fingerprint was compulsory. In an occasion whereby the fingerprint verification fails, a one-time passcodes is to be sent as an SMS to the pre-registered mobile number which would be used as a second level verification. A current study in (22) suggested a new model for improving ATM Security in Nigeria using Second-Level Authentication.

III. Methodology

The ATM second-level authentication research is simulated by a designed model encompassing of hardware and software modules to copycat a complete ATM operation as showed in Figure 1.

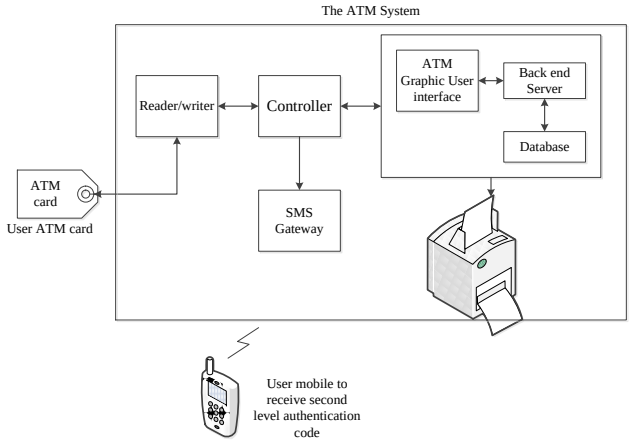


Figure 1: System Flow Diagram

The ATM card unit includes the usage of a smart card. The card can also joins to a reader with direct physical connection or with a remote contactless radio frequency interface. In this research, the physical connection card is used since unbroken communication is obligatory among the card and the reader through the operation procedure.

The card in the instance of this research is programmed via a smart card reader/writer. The smart card reader/writer can read from the card and can write to it as well. An I2C communication protocol was used to communicate with the card via the card reader. Each user details such as: account name, account number, pin number, and phone number, were written to the card for each of the user. These information are

also stored on the database for verification each time the user inserts the ATM smart card.

Smart card reader are used to read/write on smart cards. The control unit was programmed to manage the operation of the smart card reader/writer. The central control unit interfaces between the database and the user interface end. This control unit is divided into two: the software part which was implemented using Microsoft Visual studio and runs on a windows 8 All-in-One PC, the second part involves a hardware programing which runs on an Atmel microcontroller. The microcontroller monitors the user's ATM cards through a smart card reader and communicate card details to the central control unit. The microcontroller used in this research is a low-power CMOS 8-bit microcontroller based on the AVR RISC architecture.

In this research, an ultra-compact and reliable wireless module was used to achieve the SMS gateway. It has a complete Dual-band GSM/GPRS solution in a Surface Mount Technology (SMT) module which is embedded in the ATM machine system making the entire system to have a small dimensions and eventually a cost-effective solutions. Figure 2 shows the SMS gateway module connected to the database central application.

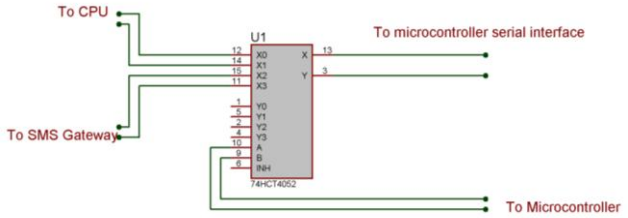


Figure 2: SMS gateway interface to CPU

The database was created using the Microsoft Access database engine. The database created is automatically stored as an SQL database hence requiring MySQL query language to access the data. A connection string is required to connect to the database. The connection string identifies a managed provider named Microsoft.ACE.OLEDB.12.0, which is an underlying database components that understands how to connect to a database and extract data from it. SQL server is one of the two most popular providers available in Visual Studio. The database contains field such as: account name, account number, account balance, and phone number etc. The frontend of the ATM App connect to the database to verify and update user account information.

The flow chart in Figure 3 shows the working operation of the system.

IV. Result and Discussion

The USB cable of the hardware unit was connected to the USB port of the computer, it was turned on and the software was lunched. If the software is launched before the hardware unit is powered on, the user receives a message "Hardware Not Connected". A green LED light comes ON to indicate that the hardware is ready. It goes OFF once a card insertion is detected. The reset button from the side of the hardware unit was pressed, then ATM application was lunched.

ATM card was inserted and a message prompted to enter the secret number, after which transactions were displayed to make a choice. Sensitive transactions involve the user to enter a second level authentication code which will be sent to user's mobile phone. A red LED flashes immediately after the SMS is sent. This totally prevent a holder of a stolen card from performing sensitive transactions such as Cash withdrawer, Changing phone number, Password change, and Setting user withdrawer limit.

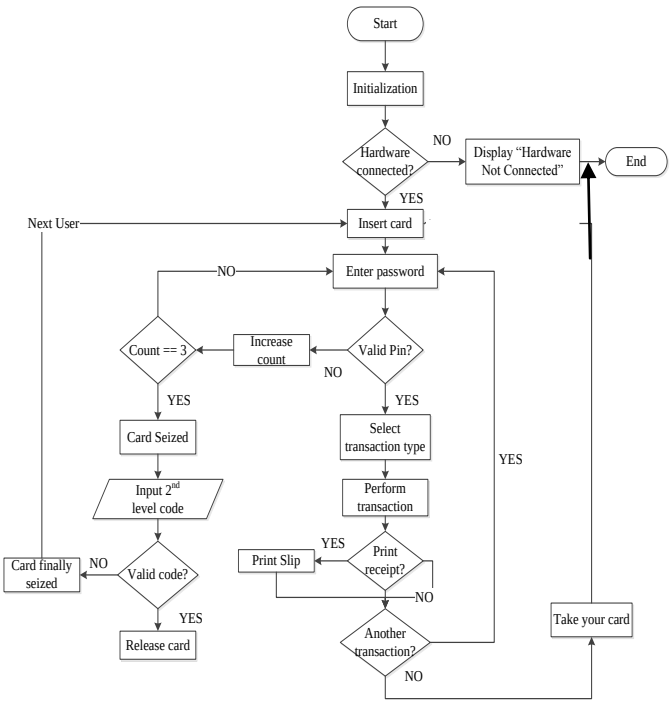


Figure 3: Operation flow chart

A. Findings

In this research, a more secured and fast way for retrieving ceased ATM card by user is proposed. After the user enters the wrong password for up to three times, the card is retained and a second level authentication code is sent to the user mobile phone. If the user is the rightful owner of the card, the card is release immediately the user key in the authentication code; otherwise, the card is finally seized. This new approach save the user the stressful process of retrieving the card using

the existing approach. This is recommended for implementation by financial institutions.

Acknowledgment

The authors acknowledge Tertiary Education Trust Fund institution-based research intervention fund (TETFUND/FUTMINNA/2014/47) through Federal University of Technology, Minna, Nigeria for sponsoring this research.

Disclaimer:

Dr. A. Ochoche is a staff of Statistics Department, Central Bank of Nigeria Head Office. The opinions expressed in this paper are those of the authors and do not necessarily reflect the position of the Central Bank of Nigeria

References

[1] R. G Jimoh, and A. N. Babatunde, "Enhanced Automated Teller Machine using Short Message Service authentication verification. World Academy of Science, Engineering and Technology," International Journal of Computer, Information Science and Engineering, Vol.8 No.1 pp. 14-17, 2014.

[2] S. A. Adepoju, and M. E. Alhassan, "Challenges of automated Teller Machine (ATM) usage and fraud occurrences in Nigeria – A case study of selected banks in Minna metropolis," Journal of Internet Banking and Commerce, Vol. 15, No. 2, pp. 1-10, 2010, Available online at: http://www.arraydev.com/commerce/JIBC/2010-08/Solomon.pdf.

[3] B. Richard, and M. Alemayehu, "Developing e-banking capabilities in a Ghanaian bank: preliminary lessons," Journal of Internet Banking and Commerce, Vol. 11, No.2, August 2006, Available online at: http://www.arraydev.com/commerce/jibc/ Accessed on: 15 September, 2014.

[4] M. I. Siddique, and S. Rehman, "Impact of Electronic crime in Indian banking sector – An Overview," International Journal of Business & Information Technology, Vol. 1, No. 2, September 2011, pp.159-164.

[5] H. B. Leow, "New Distribution Channels in banking Services," Banker's Journal Malaysia, No.110, June 1999, pp.48-56.

[6] A. A. Aliyu, and R. B. Tasmin, "Information and Communication Technology in Nigerian Banks: Analysis of Services and Consumer Reactions," In proceedings of 3rd International Conference in Business and Economic Research (3rd ICBER 2012), March 2012, pp. 150-164.

[7] O. Shoewu, and F. O. Edeko, "Outgoing call quality evaluation of GSM network services in Epe, Lagos State," American journal of scientific and industrial research, Vol. 2, No.3, 2011, pp. 409-417.

[8] S. Rosenblatt, "Two-factor authentication: What you need to know," Retrieved from: http://www.cnet.com/news/two-factor-authentication-what-you-need-to-know-faq/ Last updated on April 14, 2014. Accessed on November 23, 2014.

[9] Central Bank of Nigeria (2014), "Guidelines for Card Issuance and Usage in Nigeria," Banking and Payments System Department, May 2014. Available Online at: http://www.cenbank.org Accessed on: 25 September, 2014.

[10] O. Chima "CBN Explains Directive on Trapped ATM Cards," ThisDay Live 16 May 2014, Available online at: http://www.thisdaylive.com/articles/cbn-explains-directive-on-trapped-atm-cards-/178620/. Accessed on: 25 September, 2014.

[11] A. De Luca, M. Langheinrich, and H. Hussmann, "Towards Understanding ATM Security – A Field Study of Real World ATM Use," Retrieved from: https://cups.cs.cmu.edu/soups/2010/proceedings/a16_deluca.pdfAccessed on November 26, 2014.

[12] C. Kyle, "Biometrics: An In Depth Examination. SANS Institute Information Security Reading Room." SANS Institute 2004, Retrieved from: http://www.sans.org/reading-room/whitepapers/authentication/biometrics-in-depth-examination-1329. Accessed on November 26, 2014.

[13] N. Y. Liu, "Bio Privacy: Privacy Regulations and the Challenge of Biometrics," Taylor & Francis, 2013.

[14] S. S. Das, and S. J. Debbarma, "Designing a biometric strategy (fingerprint) measure for enhancing ATM security in Indian e-banking system," International Journal of Information and Communication Technology Research, September 2011, Vol. 1, No. 5 pp. 197-203.

[15] M. Prithika, and P. Rajalakshmi, "Card duplication and crime prevention using biometrics," IOSR Journal of Computer Engineering (IOSR-JCE), Mar. - Apr. 2013, Vol. 10, No. 1 pp. 1-7.

[16] N. Selvaraju, and G. Sekar, "A method to improve the security level of ATM banking systems using AES algorithm," International Journal of Computer Applications, June 2010, Vol. 3, No. 6, pp. 5-10.

[17] S. Ravikumar, S. Vaidyanathan, S. Thamotharan, and S. Ramakrishan, "A new business model for ATM transaction security using fingerprint recognition," International Journal of Engineering and Technology (IJET), Jun-Jul 2013, Vol. 5, No. 3, pp. 2041-2047.

[18] S. Oko, and J. Oruh, "Enhanced ATM security system using biometrics," IJCSI International Journal of Computer Science Issues, September 2012, Vol. 9, Issue 5, No. 3, pp. 352-357.

[19] V. Padmapriya, and S. Prakasam, "Enhancing ATM security using fingerprint and GSM technology," International Journal of Computer Applications, October 2013, Vol. 80, No. 16, pp. 43-46.

[20] E. Okereke, G. Ihekweaba, and F. K. Okpara, "Facial verification technology for use in ATM transactions," American Journal of Engineering Research (AJER), Vol. 02, No. 5, pp. 188-193.

[21] B. Santhi, and K. Ram Kumar, "Novel hybrid technology in ATM security using biometrics," Journal of Theoretical and Applied Information Technology, March 2012, Vol. 37, No. 2, pp.217-223.

[22] B. L. Muhammad-Bello, S. O. Ganiyu, and M. E. Alhassan, "A New Model for Enhancing ATM Security in Nigeria Using Second Level Authentication. International Journal of Science and Advanced Technology. Vol. 04, No 09, September, 2014. pp. 12-16.

Quality of Internet Service in Plateau State University, Bokkos, Nigeria

Datukun Kalamba Aristarkus
Plateau State University Bokkos, Nigeria
Corresponding Author:
kalambada@gmail.com

Sellappan Palaniappan and
Tatchanaamoorti Purnshatman
Malaysian University of Science and Technology, Malaysia

***Abstract:* this paper analysis the experience of Internet users in Plateau State University Bokkos ("PSU"). Survey was carried out on the campus, collecting data on users' experience at different levels. The goals of this research are: (1) to investigate the staff category that uses the internet most; (2) to identify the users' age groups that uses the Internet most (3); to verify the accessibility of users to the Internet and (4) to verify how often users could access the Internet. This is because learning and research is easier with Internet and could be frustrating if it is not available. Therefore, the quality of internet service, based on the satisfaction of users in Plateau State University Bokkos is analyzed in this paper.**

Keywords: Internet, Internet Service and Internet users

I. INTRODUCTION

Internet Services (IS) [9] has been key in promoting learning and research, particularly in academic environments. The provision of Internet access is basic need in any University environment due to the fact that teaching, Learning and Research are more effectively carried out when there is link to the Internet [15]. Users' satisfaction with internet service in any organization cannot be over emphasised. Many Nigerian universities today have local area networks in their libraries, finance departments, MIS Units, some faculties/departments, etc [10], yet, getting poor connectivity. The World Wide Web (www) is one of the services on the internet that offers a very large set of interlinked hypertext documents images and other resources accessed via the internet [11]. This access is possible only if the connection is available.

Advances in communication technology and the proliferation of lightweight, hand held devices with built-in, high-speed radio access are making wireless access to the Internet the common case rather than an exception. Wireless LAN installations based on IEEE 802.11 [4] technology are emerging as an attractive solution for providing network connectivity in corporations and universities, and in public places like conference venues, airports, shopping malls, and etcetera. Notwithstanding, wired LAN has been the common internet connection in beginning Universities like Plateau State University Bokkos. Even though, few cases of wireless access points are installed to create more access and convenience of access.

In this paper, we present the analysis of quality of Internet Service via respondents' (relevant university users) experiences as individuals. This information on users' experiences is collected by administering questionnaires to few users in the campus.

II. LITERATURE REVIEW

The experiences of users on the current position of Internet connection in the University are relevant in prompting technical advice to management for service upgrade. Quality of Service (QoS) is essentially an end-to-end concept of performance that occurs among users, particularly user-oriented [12]. End users on their part care only about the issues that are visible to them [14]. With the increase of various forms of high-speed (or broadband) services within today's internet, there is a new impetus to find some usable answers that allow both providers and user's to place some objective benchmark against the service offerings [16]. An End-to-End service may have a certain Quality of Service, provided for the user of a network service. It is the user that decides whether he is satisfied with the provided QoS o r not. Respondents [13] in a survey of this kind strictly consist of the relevant Internet users.

In as much as exploring issues in implementing and deploying public-area wireless networks, and exploring optimizations for improving their performance [1] is important, quality of Internet Service which directly affects the internet users is equally relevant. This is because, surviving in the information age depends on access to national and global information networks [17]. In order to evaluate and validate the techniques required for performance improvement, we consider it essential to use users' feedback on the delivery of the internet service.

Initial studies of wireless networks have explored low level error models and RF signal characteristics [3]. Installation and maintenance issues of a campus wireless network [2], user mobility in a low-bandwidth area network [7], and user behaviour and traffic characteristics in a University network [8] and/or campus [5] is important for improving Internet service delivery.

A 12-weeks survey to collected data about the Internet Service from respondents', usually the primary users of the Internet on Plateau State University Campus. Our study provides a good qualitative description of how the Internet serves the University community. In this work, we characterise user's accessibility and access consistency, taking few users within 12 weeks, filling administered questionnaire.

Other studies of wired/ wireless networks have focused more on network performance and less on user's behaviour. For example, researchers at CMU examined their campus-wide Wave LAN installation. Another study of the same campus wireless network described the issues involved in installing and maintaining a large scale wireless LAN and compared its performance to a wired LAN. A joint research effort between CMU and Berkeley [6] proposed a novel method for network measurement and evaluation applicable to wireless networks. The technique, called trace modulation, involves recording known workloads at a mobile host and using it as input to develop a model for network behaviour. This work does not involve developing a model of network behaviour, it only provide a real user's response of their experiences on the Internet as they try to surf for learning, teaching or research.

III. METHODOLOGY

In this section, we describe the methods employed in carrying out this research. This is crucial in defining the direction of the study. The research is strictly survey in nature, done by questionnaire method (particularly closed). Preparation of questionnaire based on target investigation goal, administration of the questionnaire to target users, retrieval of administered questionnaire for relevant Analysis, analysis of data collected from Internet users via questionnaire, results of Analysis and Discussion, summarizing and finally concluding on research findings constitutes the flow of this research.

See sample of questionnaire administered below:

Internet Service Respondents' Questionnaire

This questionnaire is a study of Internet use by staff and students of selected University Campuses in Nigeria, being administered by Mr. Datukun Kalamba Aristarkus in 2016 to members of the University Community. Your participation in this study is voluntary and will form part of this study and will not identify you as an individual.

Part A-Basic Questions: Please tick all that applies.

1. Sex: ☐ Male ☐ Female ☐
2. Age (years):
3. Staff: Academic ☐ Non Academic ☐

Part B- Research (Survey) Questions; choose only one question

1. Do you access the Internet Services on Campus?
   - Yes
   - No
   - Some times
2. How often do you access the Internet Service from the University campus?
   - Once a month or less
   - Once a week
   - Several times a week
   - Once day
   - Several times a day
3. Each time you access the internet in the Campus, how fast is it?
   - Very Fast
   - Partially Fast
   - Very Slow
   - Does not even open any page

Out of 150 questionnaires administered to internet users on Plateau State University campuses, only 100 were retrieved as tabulated below. This is to sample the experiences of the users about the state of Internet Service in the Campus, ascertaining the most Internet-desiring group of people on the Campuses and primarily determine user's assessment of the quality of Internet Service.

IV. RESULTS OF SURVEY

In tabulating our data retrieved from the Internet users, being the main respondents of the survey, we may use Internet users as "IU" and Internet Service as "IS". Also Plateau State University as "PSU" or "PLASU".

From Table 1 and Figure 1, we will see that more males responded to the survey than females. This implies that more males showed interest in the survey more than the females and that only 100 out of 150 participants returned the questionnaire administered. Following these facts, the users usually found on University networks may be mostly men. However, the females might have only tried to show less interest for the survey, while they actually may be consistent users of the network. If it is assumed that all the

respondents in this survey has been equally users of the network in the campus. Then, the indicated numbers can be seen as an approximated load on the University network.

Table 2 and Figure 2 describes the age groups of the respondents. We will observe that the table and the graph shows that the people who responded more to the survey were from the age of 25 to 29 years in the University. We will also notice that those of age range of 30-58 paid less attention to the survey. This implies that the age range of 25-29 utilizes the internet more, followed by that of 17-24, before that of 30-58. Having younger age participating in this survey may mean their more desire for better quality of the Internet Service on the network. Hence, may constitute much more of the loads on the network.

TABLE 1: INTERNET USERS BY GENDER

Institution	Number of Male	Number of Female	Total
PSU	60	40	100

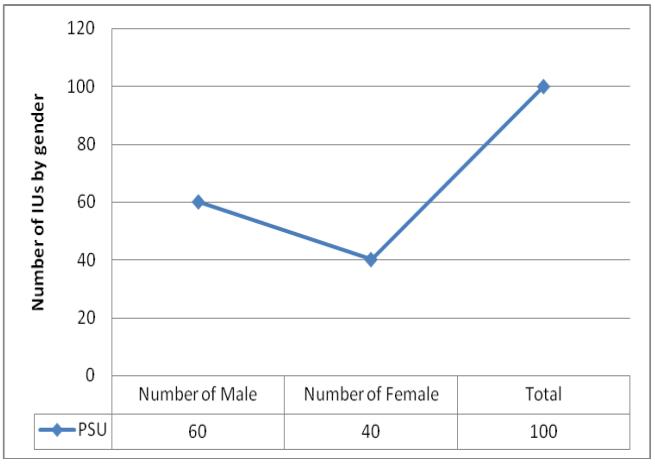


Fig 1: Graphical Description of table 1

TABLE 2: INTERNET USERS BY AGE

Institution	Age	17-24	25-29	30-58	Total
PSU		30	60	10	100

TABLE 3: CATEGORY OF INTERNET USERS

Institution	Students	Academic Staff	Non-Academic Staff	Total
PSU	60	30	10	100

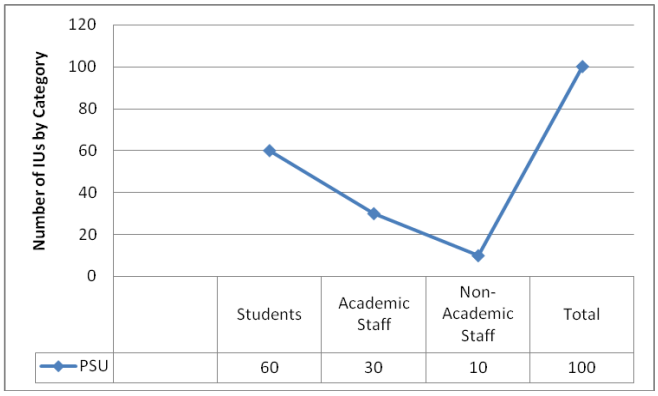


Figure 3: Graphical description of table 3

TABLE 4: ACCESSIBILITY OF INTERNET SERVICE BY IUS

Institution	Yes	No	Some Times	Total
PSU	40	10	50	100

TABLE 5: HOW OFTEN IUS ACCESSES INTERNET SERVICE

Institution	Once a month or less	Once a week	Several times a week	Once a day	Several times a day	Total
PSU	25	21	18	16	10	90

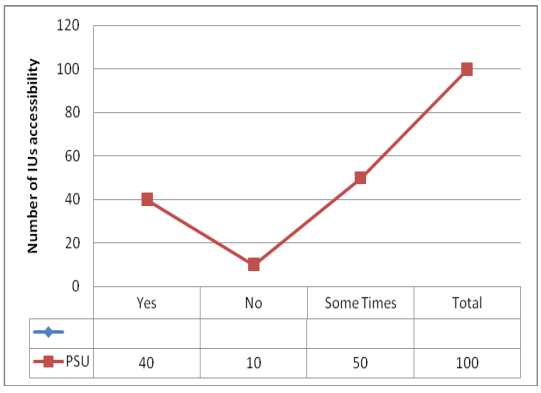


Figure 4: Graphical description of table 4

TABLE 6: INTERNET SPEED FROM IUS

Institution	Very Fast	Fast from time to time	Partially Fast	Very Slow	Does not open page	Total
PSU	15	30	15	25	5	90

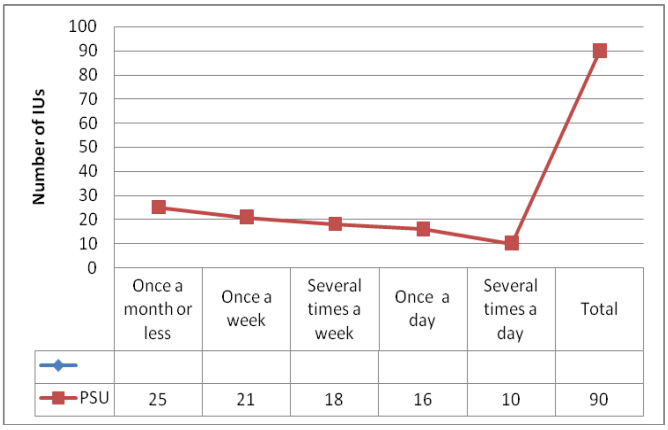


Figure 5: Graphical description of table 5

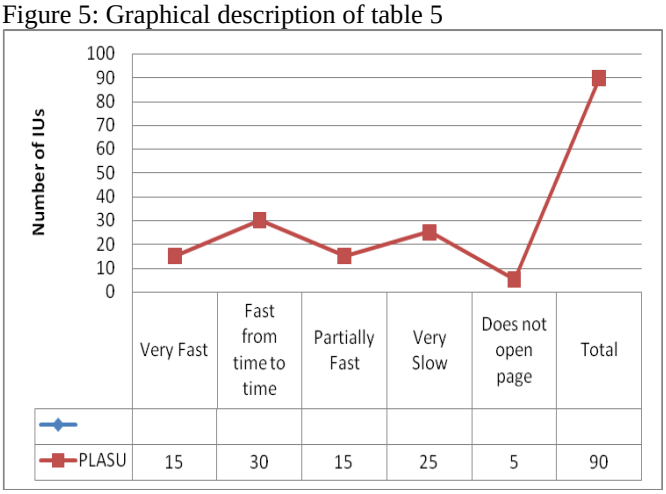


Figure 6: Graphical description of table 6

From Table 3 and Figure 3, we will observe that more students participated in this survey, followed by academic staff and lastly the non-academic staff. This shows that students and academic staff, constituting the key group that always need Internet service, utilizing the network more. By this, it entails that the main load on the networks consists of the indicated group for learning and research being their primary aim in an academic environment. Hence, in an ideal situation, the quality of Internet service of every Institution of higher learning is impactful on learning and research without reservation. We may then say that students and academic staff are more affected if the quality of Internet Service in a campus network is poor.

We can see in table 4 and figure 4 that people that admits they actually access internet are 90 out of 100 people that participated in the survey. Internet Accessibility of users is thus explained further. Out of these 90 users, 50 showed that they could only access the internet from time to time (sometimes), indicating the status of quality of Internet Service on the network. For the fact that we also have people who indicated that they could not access the internet, there is likely a form of discouraging experience in utilizing the internet, which might be as the result of poor performance of the network. Also, many were discovered to be far away from the internet centre. This means that their inability to access was due to total packet loss (from 'ping' verifications). But that of irregular access was due to inconsistent delivery, often a level of packet loss, leading to low throughput.

Considering Table 4, 10 people said they could not access the internet. So, only 90 people can tell how regular they access the internet, as shown in table 5 and figure5 , referred to as Respondents Consistency to Internet Access. We will further see in table 5 and figure 5 that only 10 out of 90 users were able to have access to the internet almost all through a day and this was discovered to be those closer to the internet centre. From the data, we are generally having irregular access to internet, referring to poor Internet service delivery status of the network. From findings, distance of nodes apart also affects the performance of network, resorting to topological and installation issues.

From Table 6 and Figure 6, only 15 users out of 90 indicated that the internet was fast, entailing that the services of the internet was not generally enjoyed by the users in the campuses in question. It was rather enjoyed by few individuals. This also shows us that the network performance was not delivering in terms of good internet speed with enough users on the network.

V. CONCLUSION

In conclusion, the quality of Internet Service in Plateau State University so far has been poor. This might be due to low bandwidth, high packet loss and/or high delivery delay. Low quality devices, cables used and probably poor network topology may also be contributing factors.

6 ACKNOWLEDGMENT

I thank all the staff and students of Plateau State University who participated in this survey, making it possible for me to carry out this research successfully. I also thank my co-authors for all their contributions in designing the necessary questionnaire and analysing the data of this research work.

REFERENCES

[1] A. Balachandran, P. Bahl, and G. Voelker "Hot-Spot Congestion Relief and User Service Guarantees in Public-area Wireless Networks" In Proceedings of WMCSA'02 – to appear. 2002

[2] B. J. Bennington and C. R. Bartel " Wireless Andrew: Experience building a high speed, Campus-Wide Wireless Data Network". In Proceedings of ACM MobiCom'97. 55–65. 1997

[3] D. Eckardt and P. Steenkiste " Measurement and Analysis of the Error Characteristics of an In-Building Wireless Network". In Proceedings of ACM SIGCOMM'96. 243–254. 1996

[4] IEEE. 802.11b/d3.0 "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification". 1999

[5] D. Kotz and K. Essien "Characterizing Usage of a Campus-wide Wireless Network". Technical Report TR2002-423, Dartmouth College. 2002

[6] B. Noble, M. Satyanarayanan, G. Nguyen, and R. Katz "Trace-based Mobile Network Emulation". In Proceedings of ACM SIGCOMM'97. 51–61. 1997

[7] D. Tang and M. Baker " Analysis of a Metropolitan-Area Wireless Network". In Proceedings of ACM MobiCom'99. 13–23. 1999

[8] D. Tang and M. Baker " Analysis of a Local-Area Wireless Network". In Proceedings of ACM MobiCom'00. 1–10. 2000

[9] Jon Dolan " Enterprise Network Proposal": Submitted to Sabah Randhawa Provost and Executive Vice President Oregon State University (OSU) and Strategic Computing Deans and Vice Provost for Research Oregon State University. 2005

[10] Nwamarah G.M. "Networking the Nigerian University System: The University of Nigeria Experience". The Seminar on Trends in The Development of ICT in the Nigerian University System. 2002

[11] Obe, O.O. and Boyinbode, O.K. "Web2.0 Platform. An inclusive open and responsive mode of governing; prospects and challenges". Conference Proceedings, Internal Conference. Managing current Global challenges; Information Technology strategies and Tool. 20. PP. 243-248, MacGlobe'09. 2009

[12] EURESCOM (). "Quality of Service: Measurement method selection". 1(1). 21. 1997

[13] Johannes Beltz "Profile of Respondents". 2005

[14] Luca Sanna Randaccio " Resources Optimization in Multimedia Communications". Department of Electrical Electronic Engineering. University of Cagliari. 2015

[15] Oluwade Bamidele, Aristarkus D. K., Ogbuju E. E. and Aregbesola M. K. (). "Technical Factors Relevant to Internet Service Provision in a Typical Nigerian University". Proceedings of the Annual Conference of the Nigerian Computer Society. ITePED 2011. 23-29. 2011

[16] Onudebelu Ugochukwu, Datukun Kalamba Aristarkus, Adewumi S. E. "Diagnosing Salem University Lokoja Network for Better Network Performance". Universal Journal of Communications and Networks. 2 (2). 40-46. 2014

[17] Igbinosa O.G. and Datukun K. Aristarkus "Evaluation of ICT Development in Nigeria using some key Indices". Pinnacle Journal Publication. 2 (1). 709-713. 2015.

Proceedings of the 1St International Conference of IEEE Nigeria Computer Chapter In collaboration with Department of Computer Science, University of Ilorin, Ilorin, Nigeria - 2016

Design and Development of a Web-Based Information System for Security Agencies

Aniji Ifesinachi Veronica[1], Onwudebelu Ugochukwu[2]

Department of Computer Science, Federal University Ndufu-Alike Ikwo (FUNAI), P.M.B. 1010, Abakaliki, Ebonyi State, Nigeria

[1]vdiamondbby@gmail.com

[2]ugochukwu.onwudebelu@fuani.edu.ng

***Abstract*—Nigeria Police Force as part of Security Agencies is known to be keeping information that is capable of securing the lives of citizens, protecting their properties, keeping peace in the country etc. And this information is being done manually in most of the various police stations which are capable of making it unsafe. Furthermore, the manual system can make processing, retrieval and referencing of information very cumbersome. Consequently, it is necessary and important to introduce and deploy computerized information system to police stations. Information Security System will help to transform manual system which is always done with pen and paper into computerized form. It is used to help in solving the problem of data redundancies, keeping security information safe, maintaining the integrity and confidentiality of information and enabling fast processing, retrieving and referencing of security information. The system has high level of security and it has several level of access which enables the uploading of information from the police officers who are the lowest in rank to the commissioner of police to the Deputy Inspector General of police and finally to the Inspector General of Police. The development tools used in the system design are Hypertext Preprocessor (PHP), Hypertext Markup Language (HTML), CSS, XAMMP Server, MYSQL and J Query.**

***Keywords*— Information, Security, Nigeria, Police force, InfoSec-NP**

I. INTRODUCTION

The term 'agency' means any executive department, military department, government parastatal, government controlled corporation, or other establishment in the executive branch of the government (including the Executive Office of the President), or any independent regulatory agency [1]. Each agency is responsible for identification of all national security systems under its ownership or control. Nigeria Police Force as part of Security Agencies is known to be keeping information that is capable of securing the lives of citizens, protecting their properties, keeping peace in the country etc.

Computer Security is defined as the prevention of, or protection against, access to information by intruders and intentional but unauthorized destruction or distortion of that information [2]. Information is vitally important to the successfully functioning of any organization. Information system occupies a vital and unique position in an organization which enables the organization to turn raw data into useful information that can be used in decision making within that organization. These information if not protected can make it vulnerable to various types of threat. Information security could be said to be a way of protecting information (data) and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction [3]. It is of great important for any organization to protect their information in order to make the information reliable, since the information stored can be referenced and retrieved whenever they are needed. Despite the high level of improvement in information technology in various organization, most Nigeria police is still in the ancient method of recording and keeping information via the manual system and these make information susceptible to damage and unauthorized access. It also hampers some defence program and delay passage of security information to the appropriate body. Consequently, an unauthorized access to security information can lead to threat to the nation at large.

The proposed Information Security System (ISS), called *InfoSec-NP*, will help the Nigeria police to eradicate the numerous problems associated with manual techniques of securing information in order to facilitate the transfer and retrieval of information. In addition, the system will enable the police to have accurate statistics to draw analysis from, on information of a case, the year a crime was committed, who investigated it, etc [4]. The data used in *InfoSec-NP* system is stored in computerized database system that will accurately and effectively record all data and information about police officers, complainant, and suspected persons. The database is a collection of interrelated data and a set of program to access those data. The database allows the generation of police, suspect and complainant report which is better storage method than paper-oriented one. The advantages of proposed system are as follow:

[1] To process the data efficiently as required.
[2] To ensure data conform to standard classifications.
[3] To reduce cost.
[4] To reduce time consumption.

[5] To increase operational efficiency.

The rest of this paper is organized as follows. Section 2 briefly reviews the relevant literature. Section 3 presents and describes the methodology used in this study. Section 4 discusses the system implementation and testing of the database. Section 5, discusses the results of our *InfoSec-NP* system. Finally, Section 6 concludes the paper by providing the study's contributions and implications.

II. Literature Review

In this section, we analyse the systems that are more related to the proposed system. Information systems appear as important instruments of investigations because they facilitate creation, storage, retrieval, transfer, and application of investigation-related information [5]. Moreover, information technologies may help produce effective use of time devoted to investigation by automating some routine investigation task [6]. Falaye, Adama, and Agemerien [7] designed a system to implements a biometric-based crime investigation system for the Nigeria Police Force (NPF). The software was designed using Visual BASIC programming language. Data entered into the database system were used for referential purposes and can be updated or modified regularly. Besides, the ability to compare fingerprints whose unique patterns assist in redundancy control was an added attribute of the system. Advantages of the system include reduction of redundancies and an inconsistency in information, ensures user defined rules to promote data integrity, enable sharing of data across all applications, and ensures proper access authorization for users. Its weaknesses are that the system was unable to generate reports and focuses more on information than crime.

Hao, Qiang and Yanfei [8] designed a mobile police information system based on web service; the system satisfies the special demands of police information systems such as security and interoperability. It enables the policemen to use various devices such as mobile phones, tablet PC, laptops etc equipped with wireless LAN or General Packet Radio Services (GPRS) to access data in the central database.

III. Methodology

The method applied in the development of this system is System/Information Engineering Model which is a phase under the Waterfall model. The reason for using this method is because it requires all part of the system. And it requires all software interacting with other parts of the system including hardware, databases and people. This method involves requirement analysis, design, coding, testing, implementation and maintenance. Requirement analysis is critical to the success of a systems or software project [9]. The requirement Analysis is the process of determining user expectations for a new or modified product. These features must be quantifiable, relevant and detailed. The requirement includes functional requirement and user requirement. The functional requirement defines what ISS is supposed to accomplish. This system, *InfoSec-NP*, will accomplish the following:

a. Administrative login.
b. Police officers update account, fill Bio-Data form and sign-out.
c. Police officers can register suspect and complainant data and statement.
d. Police officer can view and edit suspect information.
e. Police officer can upload information to the commission via e-mail.

The user requirement describes functions that are performed by the users on the system. The users of the proposed system are categorized into five levels. The requirement for this user is described using use case diagram (see Figures 1-5).

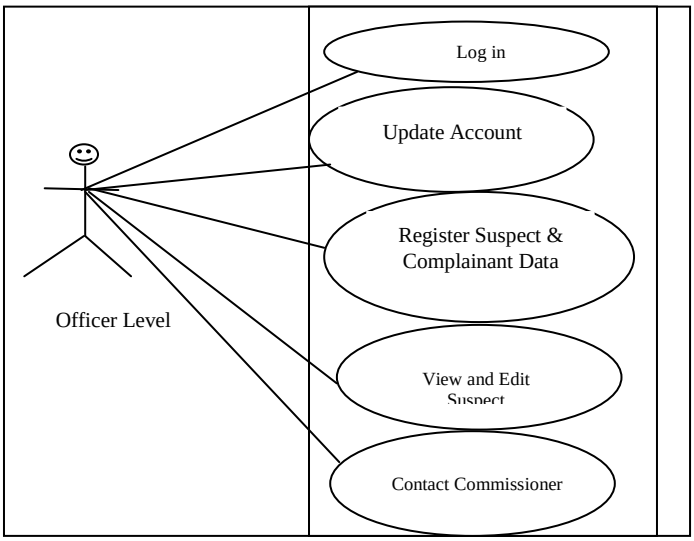


Figure 1: Use Case for Officer Level

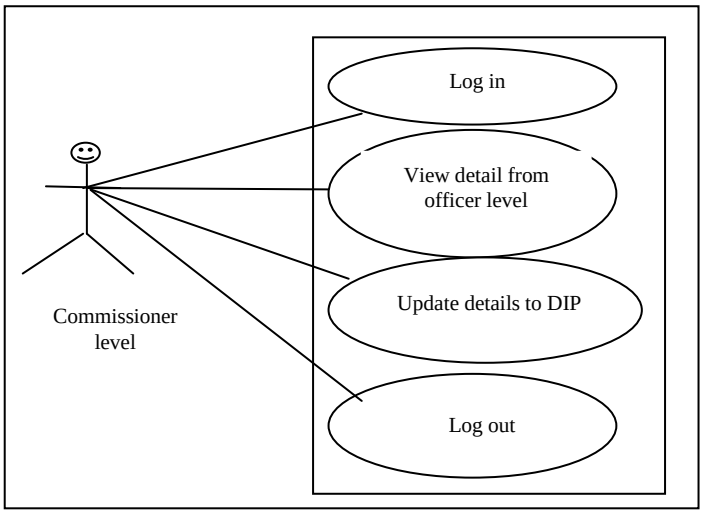


Figure 2: Use Case for Commissioner Level

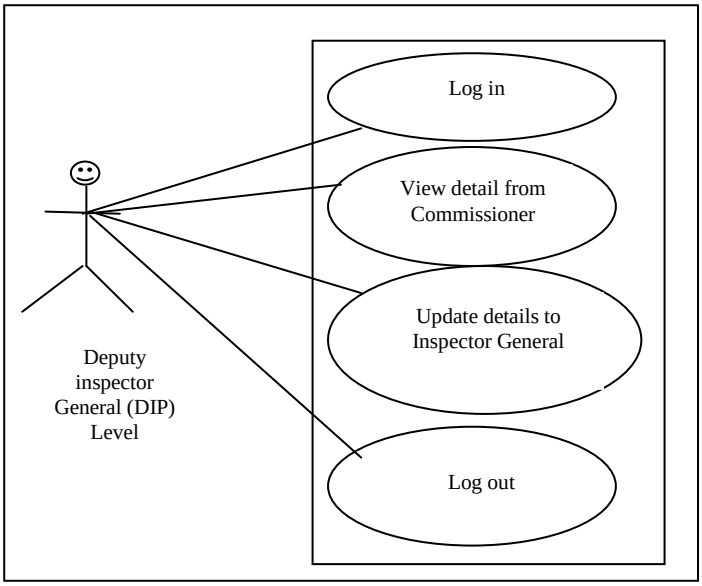


Figure 3: Use Case for DIP Level

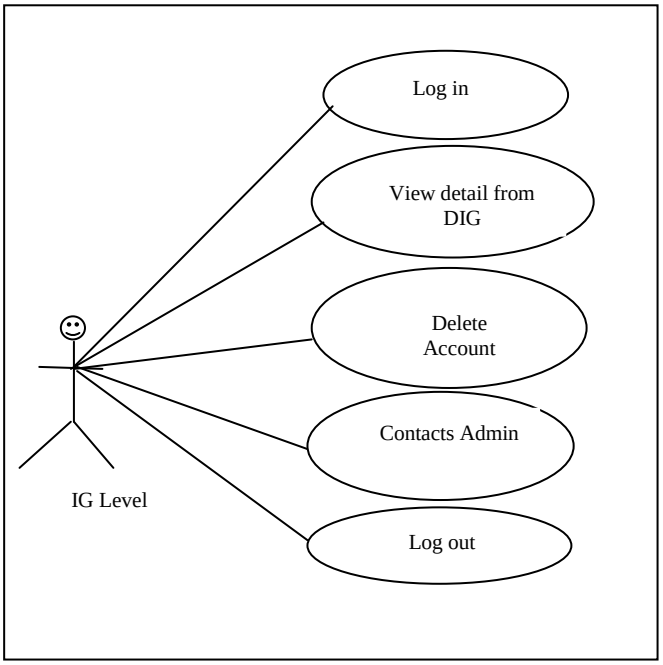


Figure 4: Use Case for IG Level

IV. System Implementation and Testing

The new system is implemented using MYSQL for the database and other Graphical User Interface based application software in order to enable the system to be user friendly and flexible. Many programming tools are considered in developing this system and some considerations are taken in order to chose the appropriate programming tools such as: Hypertext Pre-processor (PHP), a scripting language, is used for web development; Hypertext Mark-up Language (HTML) is used for describing web document, CSS for beautification of the web, XAMMP Server to run the PHP on local server, MYSQL for the database design, J Query is the cross-platform JavaScript to simplify the client-side scripting of HTML and it is used for image sliding effect and Bootstrap is a free and open-source front-end web framework for designing websites and web applications and Photoscape.

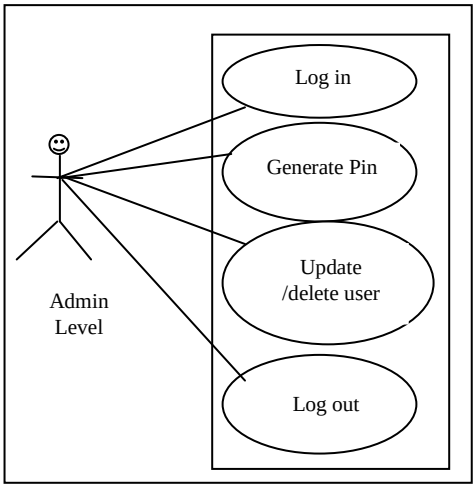


Figure 5: Use Case for Admin Level

Testing was conducted on different sections of the system for instance, the database testing. Database testing is very importance to test the database in order to achieve a database that satisfies the atomicity, consistency, Isolation, durability properties of a database management system [10]. This is so because many types of integration and implementation errors may occur in large database systems, which may cause the system to perform negatively. Table 1 shows police table, which would hold all the information concerning police officers. Table 2 is the suspect table which hold all the necessary information concerning the suspected person, his or Nickname in the case of some suspects that always go with names that are not their real names which can also help the police during investigation, date arrested, time arrested, division, name of charge room officer, name of investigating police officer, crime scene, exhibit recovered and exhibit ID. The Table 3 is the complainant table, it would hold all the necessary information about the person complained, date complained, time complained, division, name charge room officer, name of investigating police officer, statement etc.

V. Results and Discussion

After evaluating the performance of the system the results were successful as demonstrated in the outputs in *Figures 6-8*.

i. The police login was successful.
ii. The police bio-data was successfully captured (*see Figure 6*).

iii. The police was able to key in complainant bio-data and complain (*see Figure 7*).
iv. The police was able to key in suspect bio-data and statement (*see Figure 8*).
v. The police was able to update their few profiles.
vi. The police contacted the commissioner successfully using email address.

TABLE I: POLICE TABLE

police_data

Column	Type	Null	Default	Comments	MIME
System_ID	int(11)	No			
RegDate	timestamp	No	CURRENT_TIMESTAMP		
Username	varchar(20)	No			
Password	varchar(20)	No			
User_level	varchar(15)	No	user		
Zone	varchar(10)	No			
State	text	No			
Area	text	No			
Division	varchar(20)	No			
Rank	varchar(20)	No			
Surname	text	No			
Firstname	text	No			
Middle_Name	text	No			
Nick_Name	varchar(20)	No			
Person_Gender	text	No			
Date_Of_Birth	varchar(15)	No			
Address	varchar(50)	No			
Force_No	varchar(20)	No			
Date_Recruited	date	No			
Mobile_No	varchar(15)	No			
Drivers_Licence_No	varchar(50)	No			
Nationality	text	No			
Tribe	text	No			
Religion	text	No			
State_of_Origin	text	No			
LGA	text	No			
Email	varchar(50)	No			
Passport	varchar(50)	No			

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	System_ID	0	A	No	

Print

TABLE 2: SUSPECT TABLE

criminal_data

Column	Type	Null	Default	Comments	MIME
ID	int(11)	No			
RegDate	timestamp	No	CURRENT_TIMESTAMP		
Surname	text	No			
Firstname	text	No			
Middle_Name	text	No			
Nick_Name	varchar(20)	No			
Person_Gender	text	No			
Address	varchar(50)	No			
Occupation	text	No			
Age	varchar(10)	No			
Religion	text	No			
Tribe	text	No			
Nationality	text	No			
Date_Arrested	date	No			
Time_Arrested	time	No			
Division	varchar(50)	No			
CRO	varchar(20)	No			
IPO_Incharge	varchar(20)	No			
Caution	text	No			
Statement	text	No			
Crime_Scene	varchar(50)	No			
Exhibit_Recoverd	varchar(50)	No			
Exhibit_ID	varchar(50)	No			
Accused_Signature	varchar(50)	No			
IPO_Signature	varchar(20)	No			
Date_Released	date	No			
Time_Released	time	No			
Passport	varchar(20)	No			

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	ID	0	A	No	

Print

TABLE 3: COMPLAINANT TABLE

complaint_table

Column	Type	Null	Default	Comments	MIME
ID	int(11)	No			
RegDate	timestamp	No	CURRENT_TIMESTAMP		
Surname	text	No			
Firstname	text	No			
Middle_Name	text	No			
Address	varchar(50)	No			
Occupation	varchar(30)	No			
Age	varchar(10)	No			
Sex	text	No			
Religion	text	No			
Tribe	text	No			
Nationality	varchar(30)	No			
Mobile	varchar(20)	No			
Date_Complained	date	No			
Time_Complained	time	No			
Division	varchar(50)	No			
CRO	varchar(50)	No			
IPO_Incharge	varchar(50)	No			
Statement	varchar(500)	No			
Complaint_Signature	varchar(20)	No			
IPO_signature	varchar(20)	No			

Indexes

Keyname	Type	Unique	Packed	Column	Cardinality	Collation	Null	Comment
PRIMARY	BTREE	Yes	No	ID	0	A	No	

Print

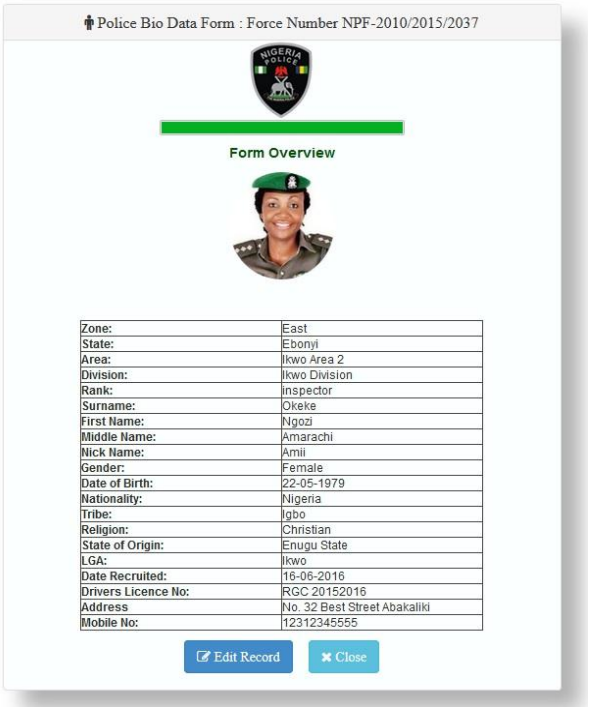


Figure 6: Police Bio Data Form

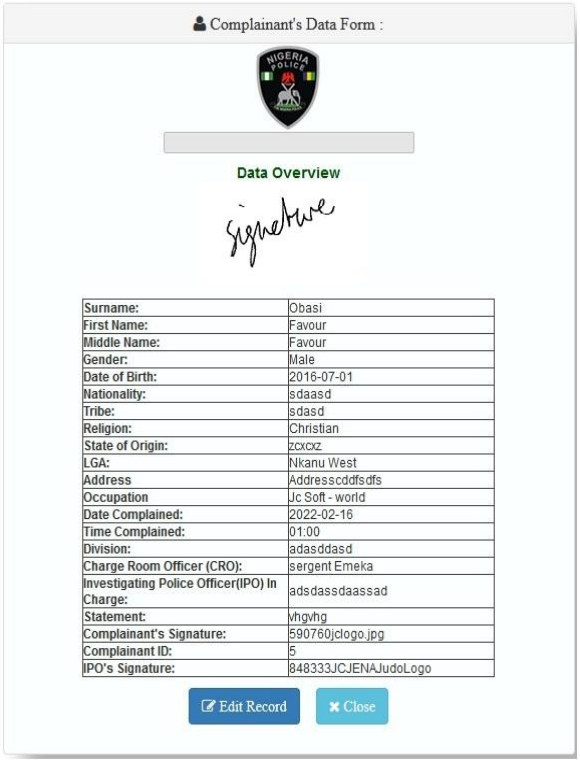


Figure 7: Complainant's Data Form

In the course of developing the *InfoSec-NP* system, the major challenge encountered was the gathering of all the necessary requirements from the Nigeria police force. Closely related systems also helped in extracting system requirements. In addition, the system also was able to generate police report, suspect report and complainant report.

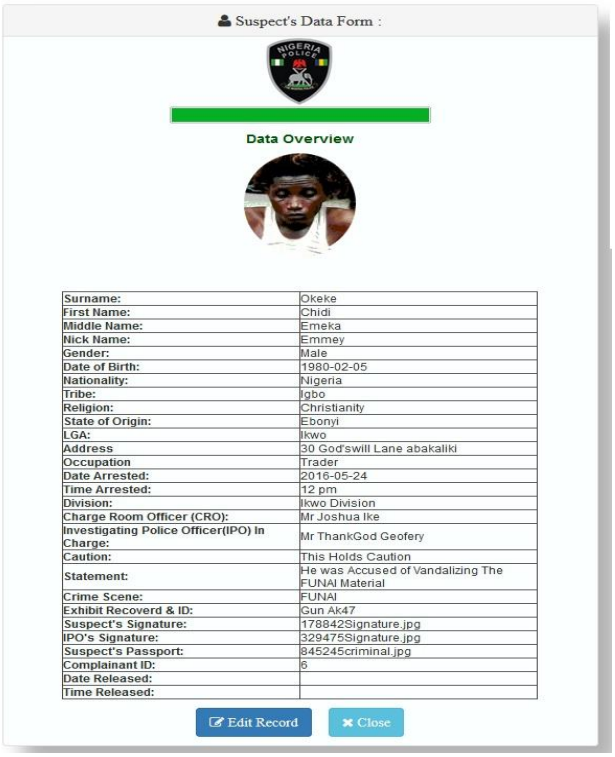


Figure 8: Suspect's Data Form

VI. CONCLUSION

The need for a computerized platform for keeping information in Nigeria police cannot be overemphasized. A system that provides the Nigerian police with security information regarding the various activities they had carried in the recent past, will go a long way in helping to keep information safe and secured; and this will also help in keeping record of new cases of crimes within a given locality. The web-based information system for security agencies enhances proper and efficient processing and retrieving of information by the Nigeria Police Force, thereby, helping in making informed decisions and improving reliability, thus, improving security agencies operations in keeping information safe. This will results in low rate of information lost. The automated system that has been developed will facilitate the keeping of records of suspect for future references. If adopted, it will be difficult for police officer to tamper with any

information being keyed into it. And there will be no case of lost of case file or the release of suspects with impunity.

REFERENCES

[1] W. C. Barker. Information Security: Guideline for Identifying an Information System as a National Security System. 2003

[2] S. Ross (1999) Unix System Security Tools. Computer Security a Practical Definition, http://www.albion.com/security/intro-4.html

[3] CNSS (2010) Committee on National Security System, National Information Assurance (IA) Glossary CNSS instruction, No. 4009, 26 April 2010, Retrieved in 15th March 2016 from http://en.m.wikipedia.org/wiki/information_security.

[4] N. Peter. Crime Mapping for Effective Policing, 2010, Retrieved on 24/05/2016 from http://www.gbooza.com/group/crime/forum/topics/crime-mapping-for-effective#axzz1NGw1881w

[5] D. Gottschalk. Knowledge Management Systems in Law-enforcement: Technologies and Techniques, idea group pub, Hershey, PA. 2007

[6] J. E. Eck Solving Crimes: The Investigation of Burglary and Robbery, Police Executive Research Forum, Washington, DC. 1983

[7] Falaye, A., Adama, N. & Agemerien, F. Design and Implementation of Crime Investigation System using Biometric Approach (Nigeria Police Force), The Pacific journal of science and Technology. 2013

[8] Y. Hao, F. Qiang & F. Yanfei. Mobile Police Information System Based on Web Services, Tsinghua Science and Technology, 11 (1), pp. 1-7, 2006

[9] A. Alain and J. W. Moore Chapter 2: Software Requirements. Guide to the software requirement engineering body of knowledge, Bourque, P. & Dupuis, R. editor (2004 ed.), Los Alamilos, CA : IEE computer society press. ISBN 0-7695-2330-7, 2005.

[10] H. Korth. Database System Concepts, Sixth Edition. Macgraw-Hill January 28, 2010.