THE SECONDARY USE OF IDENTIFIABLE PHYSIOLOGICAL DATA FOR HEALTH RESEARCH PURPOSES

Yvonne Choi 1, Carolyn McGregor 2

1Faculty of Health Sciences, University of Ontario Institute of Technology, Oshawa, Canada
2Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, Canada

yvonne.choi@uoit.ca, carolyn.mcgregor@uoit.ca

This paper provides an overview of the essential role that the secondary usage of personal health data enables for purposes of expanding current knowledge and understanding of health care and its delivery. Although the use of health data for secondary purposes such as research has the potential to benefit society as a whole, such data are often sensitive in nature and/or personally identifiable, raising privacy concerns that it can also be misused. For example, physiological data (e.g. electrocardiogram [ECG] data) holds the potential to reveal more information about an individual than one may realize.

I. INTRODUCTION

Since its origin as a science in 1948, informatics has successfully proven itself to be a valuable discipline as it allows for the linkage of any and all kinds of data, knowledge and processes [1]. Building on current knowledge and technology, the analysis of big data and data analytics holds promise for new discoveries to improve healthcare outcomes.

Large, electronic databases serve as valuable resources in the area of health and health services research. Secondary usage of health data is defined as the use of personal health information collected for purposes unrelated to the initial purpose of providing direct delivery of health care to the individual that the information concerns. This includes activities such as research, analysis, quality and safety measurement, payment, provider accreditation and commercial activities [2]-[3]. Such data may be classified as identifiable.

While much research can indeed be performed without the need for personally identifiable data (i.e. by utilizing anonymised data), it is acknowledged that the use of such data may be required to address certain research questions and/or for methodological purposes (e.g. linkage within a database or between databases, ensuring of meaningful comparison and completeness of recruitment, investigation of social factors and for assessing the applicability of research findings). In the study of health and health care, primary research areas that depend heavily on the secondary usage of identifiable health data include the: 1) natural history and development of diseases; 2) etiology of diseases; 3) evaluation of health care interventions; 4) equity of health care; 5) identification of health care services trends; and 6) to support primary research [2].

This paper highlights the importance of using identifiable physiological data for health research purposes as well as the urgent need to raise public awareness and development of new approaches in balancing legitimate privacy concerns related with the secondary use of physiological data, such as ECG data.

II. TECHNICAL INFORMATION

A. ARTEMIS PLATFORM AND PHYSIOLOGICAL DATA COLLECTION

Artemis is a platform for online health analytics that allows for concurrent multi-patient, multi-diagnosis and multi-stream temporal analysis of complex, high-frequency physiological data streams in real-time for purposes of clinical management and research. By comparing the analytical results that are gathered in the platform with current treatment practices, new patterns in real-time physiological data can be identified thus enabling detection and prevention of various health conditions before clinical symptoms are visible. Artemis captures ECG data as well as derived signals from the ECG for purposes of breath detection (e.g. heart rate, respiration rate and chest impedance). When available, other signals captured provide information such as blood oxygen saturation levels in addition to diastolic, systolic and mean blood pressure [4].

B. BIOMETRICS

Biometric data is personal information derived from an individual to determine or verify one’s identity [5]. The term “biometrics” may refer to quantifiable characteristics or the automated methods that utilize the aforementioned characteristics to identify or confirm one’s identity [6]. Any human behavioural and/or physiological characteristic has the potential to be utilized as a biometric identifier provided it satisfies the criteria of universality, distinctiveness, acceptability collectability, performance permanence and circumvention [7]. Physiological biometrics are based on data derived from a direct measurement from a part of the human body [5].
Studies have revealed that physiological data such as the ECG trace contains features that are unique to an individual [8]. Majority of existing biometric identification approaches rely on amplitude and temporal distances between detected fiducial points from the ECG [9]. The fiducial points exploit an individual’s unique physiology as they have fixed positions relative to one’s heartbeat and contain stable features that are characteristic to them and are unaffected by the individual’s anxiety state or sensor location [8]. More recently, an ECG biometric identification method without fiducial detection has been proposed [10] that can potentially eliminate the impracticalities of relying on a fiducial detector and difficulties related with accurate fiducial detection [9].

While the use of ECG as a biometric can prove to be challenging, the long-term stability of relying on ECG data was examined in a feasibility study that assessed the utilization of ECG characteristics for human verification and identification purposes. Short ECG recordings were taken repetitively from subjects over the course of several months to years. Using the heart vector and a simple distance measure, it was found for verification purposes, the equal error rate that was achieved was 2.8% while for identification purposes, 98.1% of the individuals were correctly identified, suggesting that ECG can be beneficial as an add-on to existing biometrical systems [11].

C. DATA SHARING AND PRIVACY CONCERN

The increased interest in e-science and its associated technologies is the consequential result of multiple factors including the urgent need to manage the ever-growing quantities of complex data, the pressure to maximize return on research investments (i.e. funds), for purposes of ensuring methodological soundness and addressing of research questions that can only be answered via the analysis of large and/or merging of otherwise disparate datasets [12]. But as technologies continue to be increasingly interconnected in the big data era, collections of personal information will become limitless. The dependence on compliance with legislation alone to ensure the protection of privacy can no longer suffice [13]. Since physiological data can be linked back to the specific individual from whom the data was captured from, it can potentially be abused or misused for purposes that one did not consent to thus infringing on one’s privacy. As privacy is experienced on a personal level, the notion of privacy has different connotations to each individual. Despite the lack of a coherent “definition”, privacy is nonetheless considered to be valuable as it promotes fundamental values such as personal autonomy, individuality, dignity, respect and worth as human beings while others describe privacy as a basic human good or right that has a value that is intrinsic [15].

III. CONCLUSION

As with any technology, biometrics are defined by their usage and the systemic and operational controls that are put into place. For that reason, biometrics can be deployed in a privacy-invasive, privacy-protective or privacy-neutral fashion. If properly employed, biometrics can be seen as a privacy-enhancing technology as biometric data can protect other sensitive personal information [7].

IV. REFERENCES