Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230092

An Enhanced Standardization and Qualification Mechanism for Heterogeneous Healthcare Data

George MANIAS[a,1], Ainhoa AZQUETA-ALZÚAZ[b], Andrea DAMIANI[c], Eshita DHAR[d], Eleftheria KOUREMENOU[a], Marta PATINO-MARTÍNEZ[b], Mariachiara SAVINO[c], Syed Abdul SHABBIR[d] and Dimosthenis KYRIAZIS[a]
[a] *University of Piraeus, Piraeus, Greece*
[b] *Universidad Politécnica de Madrid, Madrid, Spain*
[c] *Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy*
[d] *Taipei Medical University, Taipei, Taiwan*

Abstract. Given the challenge that healthcare related data are being obtained from various sources and in divergent formats there is an emerging need for providing improved and automated techniques and technologies that perform qualification and standardization of these data. The approach presented in this paper introduces a novel mechanism for the cleaning, qualification, and standardization of the collected primary and secondary data types. The latter is realized through the design and implementation of three (3) integrated subcomponents, the Data Cleaner, the Data Qualifier, and the Data Harmonizer that are further evaluated by performing data cleaning, qualification, and harmonization on top of data related to Pancreatic Cancer to further develop enhanced personalized risk assessment and recommendations to individuals.

Keywords. Data Standardization, Data Qualification, Healthcare Analytics

1. Introduction

Nowadays, the healthcare domain faces various challenges related to the diversity and variety of data, their huge volume, and their high distribution. While the massive investments by the healthcare industry into new technologies and the rapid growth in the usage of cloud and mobile computing, medical devices, IoT, and Artificial Intelligence (AI) lead to the increasing need for the design and utilization of enhanced and state-of-the-art healthcare analytics solutions [1]. Thus, the successful cleaning, quality assurance, as well as interpretation and harmonization of the heterogeneous data can provide more precise and personalized prevention and intervention measures, higher experience for patients' health monitoring, and personalized decision support. To address all these challenges this paper introduces a novel mechanism for the cleaning, qualification, and standardization of the collected data. The next section states state-of-the-art standardization and quality assurance techniques, introduces the novel mechanism and its subcomponents, and concludes this research work.

[1] Corresponding Author: George Manias, 80, M. Karaoli & A. Dimitriou St., 18534 Piraeus, Greece; E-mail: gmanias@unipi.gr.

2. Standardization and Qualification Methods and Proposed Approach

Recently, an extensive literature review was performed to existing Clinical Prediction Models (CPMs) [2] examining the different data cleaning approaches that are applied on them. The results indicated that all of them are described by diversity, inconsistency, and lack of reported details in how missing data are handled. As concerns the quality assurance of the healthcare data, different qualitative and quantitative measures and methods are proposed to assess the quality of the data [1]. In addition, the ever-increasing usage of EHRs during the last decades enables more effective and efficient data sharing. However, the use of different clinical content and terminology standards across different healthcare organizations is quite common and thus the lack of a wider standardization for data sharing continues to be a major impediment in achieving true data interoperability in the healthcare domain [3]. Under the scopes of this research work, the proposed mechanism seeks to address the aforementioned lacks and challenges in both the qualification and standardization of healthcare data. To this end, the proposed mechanism exploits three (3) processing phases, the cleaning, the qualification, and the harmonization of the data. These phases are realized through the design and implementation of three (3) integrated and automated subcomponents, i.e., the Data Cleaner, the Data Qualifier, and the Data Harmonizer as depicted in Figure 1.

Figure 1. Standardization and Qualification Mechanism

The proposed mechanism assures the incoming data's accuracy, integrity, and quality, while it also provides a decision whether a connected data source will be considered as reliable or not. On top of this, it enhances the interoperability of data through automated standardization techniques that are applied on the cleaned and qualified data. Through its utilization data are presented in a consistent manner irrespective of the data source. The proposed mechanism will be further applied and evaluated in the context of a novel personalized-healthcare framework as realized in the context of the iHELP project [4].

Acknowledgment

The research leading to the results presented in this paper has received funding from the European Union's funded Project iHELP under grant agreement no 101017441.

References

[1] Cichy C, Rass S. An overview of data quality frameworks. IEEE Access. 2019 Feb 15;7:24634-48.
[2] Tsvetanova A et al. Missing data was handled inconsistently in UK prediction models: a review of method used. Journal of Clinical Epidemiology; 2021; 140: 149-158.
[3] Kubben P, Dumontier M, Dekker A. Fundamentals of clinical data science; 2019.
[4] Manias G, et al. iHELP: Personalised Health Monitoring and Decision Support Based on Artificial Intelligence and Holistic Health Records. In 2021 IEEE Symposium on Computers and Communications (ISCC); 2021; 1-8; IEEE.