1st Serbian International Conference on Applied Artificial Intelligence (SICAAI)
Kragujevac, Serbia, May 19-20, 2022

HYBRID MEDICAL RECORDS SYSTEM FOR ARTIFICIAL INTELLIGENCE: ESSENCE AND CHALLENGES

Miroslav Bojović[1], Dragan Bojić[1], Nenad Korolija[1], Veljko Milutinović[2]

[1]Department of Computer Technology and Information Theory, School of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11000, Belgrade, Serbia
e-mail: mbojovic@etf.rs, bojic@etf.rs, nenadko@etf.rs
[2] Department of Computer Science, University of Indiana in Bloomington, Luddy Hall 2062, 700 N. Woodlawn Avenue Bloomington, IN 47408, Indiana, United States
e-mail: vmilutin@iu.edu

Abstract:

Healthcare systems can benefit from artificial intelligence. Compared to many other domains in which artificial intelligence is exploited, these systems require extra caution. The distinguishing difference is that the medical personnel cannot rely solely on the recommendation given by the system, without any clue about the reasoning behind.

Many efforts have been made to exploit potentials of artificial intelligence to transform how care is delivered. Results include using robots for communication purposes, extracting data for further processing using sensors, and helping medical personnel in the decision process. However, much of the research is oriented towards using AI for extracting hidden data from vast of available medical data.

The focus of this work is on defining the new structure of Electronic Health Records, connecting genomics data, existing knowledge of medical personnel, and data streamed from sensors to form an extensible big data architecture appropriate for AI. By combining the benefits of the accuracy of deep-learning algorithms with ages of human experience, resulting in automated healthcare recommendations along with factors that lead to making the decision that are understandable to physicians. The proposed AI approach has the encoded clinical guidelines and protocols that serve as a starting point augmented by AI models that learn from data and discover hidden data. The proposed AI Healthcare system incorporates two concepts starting with rules-based approaches, but increasingly introduces the capacity to learn by training a specific task on large datasets and demonstrate the distinctive properties of AI, which also open new routes to delivering better, faster and more cost-effective care, and may have a greater focus on prevention and promoting wellness.

Keywords: artificial intelligence, healthcare, electronic health records

1. Introduction

Advances in medicine, supported by innovation in technology, are accelerated dramatically in recent years. By the 1950s, medical knowledge had doubled in about 50 years. In 2020, the volume of medical knowledge will double in 73 days [1]. This evolution has significantly risen average life over the past century from less than 50 years to 78.9 years for the USA and to 80.9 years for EU [2]. By 2050, 1 in 6 people will be over the age of 65, in Europe and North America, this will be 1 in 4. This demographic shift, combined with rapid urbanization, modernization, globalization and accompanying changes in risk factors and lifestyles, means chronic conditions will be more common, and an increasingly comorbid population’s demand for healthcare will increase [3].

In recent decades, we are witnessing increased usage of artificial intelligence in various domains of science and technology. Healthcare system is not an exception. Which artificial intelligence in healthcare, a medical institution can significantly improve self-care, prevention, overall wellness, triage and diagnosis, clinical decision support, care delivery and management, etc. At the same time, contemporary healthcare systems face growing demand for their services along with rising costs.

Following chapters present the problem definition, current trends, and proposes a new uniform model for processing data from multiple sources to be used for AI. Analysis of the proposed experimental model is followed by the conclusion. This work is described using the proposed method for presenting research results [4].

2. Problem Definition

This section presents the definition of the problem that healthcare systems that exploit artificial intelligence have to solve.

As it is the case with any decisions that might affect someone's life and wellbeing, the requirements from artificial intelligence systems are that:

- The reasoning has to be understandable.
- The decision making process can be based on deduction and known facts.
- The reasoning has to be based on available existing data from multiple resources.

This pepr is related to modelling the data from various resources and applying artificial intelligence in such a manner that the decision steps are understandable by medical care professionals.

3. Existing Technologies for Healthcare Systems

Computers are utilized in healthcare for decades, ranging from relational database systems [5], over hardware acceleration of most commonly used operations [6] to complex simmulations of fluid dynamics [7]. Many efforts have been made to exploit potentials of Artificial Intelligence to transform how care is delivered. Some of them are related to using robots for communication purposes and stimulating patients for activity [8, 9]. Others use sensors to extract data that will undergo the automated reasoning process [10, 11]. Effort has been made to exploit AI for advising medical personel [12].

While deep learning techniques produce state-of-the-art performance on a variety of tasks, one of its main criticisms is that the resulting models are difficult to naturally interpret. In this regard, many deep learning frameworks are often referred to as "black boxes", where only the input and output predictions convey meaning to a human observer. Since correct clinical decision-making can be the difference between life and death, many practitioners must be able to understand and trust the predictions and recommendations made by deep learning systems. There are authors attempts and approaches [13] to make clinical deep learning more interpretable such as: Maximum Activation: A popular tactic in the image processing community is to examine the types of inputs that result in the maximum activation of each of a model's hidden units. This represents an attempt to examine what exactly the model has learned, and if it can be used to assign importance to the raw input features; Constraints: Others have imposed training constraints specifically aimed at increasing the interpretability of deep models. For example, Choi et al.'s Med2Vec framework [14] for learning concept and patient visit representations uses a non-negativity constraint enforced upon the learned code representations; Qualitative Clustering: In the case of EHR concept representation and phenotype studies, some studies point to a more indirect notion of interpretability by examining natural clusters of the resulting vectorized representations. This is most commonly performed using a visualization technique known as t-Distributed Stochastic Neighbor Embedding (t-SNE), a method for plotting pairwise similarities between high-dimensional data points in two dimensions [15]; Mimic Learning: Che et al. [16], [17] first train a deep neural network on raw patient data with associated class labels, which produces a vector of class probabilities for each sample. They train an additional gradient boosting tree (GBT) model on the raw patient data, but instead use the deep network's probability prediction as the target label. Since GBTs are interpretable linear models, they are able to assign feature importance to the raw input features while harnessing the power of deep networks. The approaches given and known from the open references cannot be recognized as full naturally interpretable, and are not widely accepted.

4. Proposed Solution for Healthcare Systems

The proposed system defines a new structure of Electronic Health Record (EHR) that can connect longitudinal data providing insights across episodes of treatment and settings of care, and incorporating new types of data from wearables, mobiles, NL data, sensors and genomics and other omics data. The digital revolution in healthcare provides new ways to both collect high-quality data from each patient and connect it to data from large pools of patients for analysis with AI algorithms.

Another innovative contribution of this project is an extensible big data architecture appropriate for this AI concept. The big data technologies make it possible to collect huge volumes and wide spectrum of data, molecular information generated from next-generation sequencing, data from wearable devices and mobile apps, and novel clinical examinations and experiment.

The process of extracting from big data can be broken down into five stages [18]. These five stages form the two main sub-processes: data management and data analytics. Data management involves processes and supporting technologies to acquire and store data and to prepare and retrieve it for analysis: Acquisition and Recording; Extraction, Cleaning and Annotation; and Integration, Aggregation and Representation. Analytics, on the other hand, refers to techniques used to analyse and acquire intelligence from big data. The proposed extensible big data architecture consists of Data Acquisition, Data Semantic and AI Data Analytics modules. The Data Acquisition module is responsible to collect data from various sources. This component may be combination of HDFS, NoSQL such as MongoDB and SQL database. The Data Semantic module is mapping heterogeneous databases into common structure and semantics. The Web Ontology Language with XML syntax can be used as standard interchange format regarding ontology. The thorough analysis through the design and development of these modules is expecting to suggest the most suitable solution. The AI Data Analytics module is responsible for Extraction, Cleaning and Annotation, Integration, Aggregation and Representation, and is based on AI mechanisms proposed through the Main AI module proposal. The AI Data Analytics module compares every processed data to predefined user's threshold. If the value of a particular data exceeds alarming threshold value, it will be stored while an emergency alert is generated to the Main AI module. The Main AI module decides if it is value of data for final emergency alert, or regular storage in EHRs. The proposed extensible big data architecture is capable of batch and stream processing and be based on available frameworks. Various available frameworks will be analysed. For batch processing mode Hadoop allows distributed big data on a cluster of machines but may not be appropriate for stream processing. When stream processing is required, Storm, S4, Apache Spark, Apache Flink may be a suitable choice. Apache Spark may be serious candidate. It is open source unified engine for distributed data processing that includes higher-level libraries for supporting SQL queries (Spark SQL) that restore data from many sources and manipulate them using SQL, streaming data (Spark Streaming), iterative machine learning algorithms through library mechanism (MLlib), provides efficient algorithms with high speed, structured data analysis using Hive, and graph processing based on GraphX. Further information extraction and processing from EHRs requires specialized toolsets for Natural Language Processing, Image Analytics (Visualization Toolkit, GIMIAS, Elastix, MITK), Machine Learning (Tensorflow, Keras, Theano, Torch, Caffe...) and analytics for "Omics" data (SparkSeq, SAMQA, Distmap, Hydra...).

The solution proposed in this manuscript is based on the one previously described in a book chapter [19]. On top of the previous work, we've modeled a wrapper classes for parsing the data from various resources in a uniform manner, accompanied by a function that periodically adapts the artificial intelligence knowledge based on new data streamed over sensors and is received by medical professionals. Implementation guidelines described in [19] are extended by modelling a process that automatically feeds data into the artificial intelligence model from various sources. The model is extendable and supports adding new types of input data to be combined with existing ones.

5. Analysis and Comparison

Compared to the existing solutions, the proposed one is more general. It not only exploits deep learning techniques for discovering hidden knowledge, but the recommendations to medical professionals are given in a set of directives based on commonly accepted reasoning, and the data feed from various resources including sensors is fed to the artificial intelligence model in a real time.

The additional innovative contribution of the proposed system is a combination of two AI concepts, where encoding clinical guidelines and/or existing clinical protocols, as in rules-based systems, provides a starting point, which then can be augmented by deep learning models that learn from data. The first AI concept makes that the decision explainable and the factors that are important to the algorithm's conclusion are visible which rise the confidence of the physicians and other practitioners. The second AI concept demonstrates the distinctive properties of AI, which also open new routes to delivering better, faster and more cost-effective care, and may have a greater focus on prevention and promoting wellness.

Our goal is not that the proposed system replaces physicians and other professionals. Physicians and other professionals who use AI will replace others who do not use AI. AI needs to serve as a decision-support tool. In the end, it is the decision of a doctor. AI is not going to substitute doctors in the foreseeable future. AI will be able to do some tasks that take a lot of time, need interconnection of many complex data, use mathematical formulas to determine the correct dosage of drugs, use numerous practical cases for data mining and analytics, what can free up physician time for concentration on diagnostic thinking and talking to patients.

Consequently, the proposed system has the potential to transform healthcare organizations and healthcare services, and meet the current and future society needs not only in Serbia, but in many other countries.

6. Conclusions

Healthcare systems can benefit from artificial intelligence. Unlike many other fields of use, healthcare systems cannot just benefit from the data obtained by automated reasoning but causing serious problems to patients are possible due to the potential reasoning on the data that is not representative. This paper presents a uniform model for feeding the artificial intelligence system with data from various sources. It also suggests that artificial intelligence reasoning has to be understandable to medical care professionals, as they are responsible for making decisions and have to be aware of necessary consequences for making each decision.

References

[1] Mesko B, A Guide to Artificial Intelligence in Healthcare, The Medical Futurist, 2019.

[2] Densen P, Challenges and opportunities facing medical education, Transactions of the American Clinical and Climatological Association, 122, 2011.

[3] Life expectancy at birth in the EU, Eurostat, July 2019. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20190725-1. 2019 Revision of World Population Prospects, UN.

[4] Milutinovic V, "The best method for presentation of research results," IEEE TCCA Newsletter, 1-6, 1996.
[5] Bojovic M and Bojic D, MobilePDR: A mobile medical information system featuring update via Internet, IEEE Transactions on information Technology in Biomedicine, 9(1):1-3, 2005.
[6] Velasevic DM and Bojovic MM, Hardware support to operations of relational algebra, Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences, Architecture Track, IEEE Computer Society, 1:191-192, January 1989.
[7] Korolija N, Djukic T, Milutinovic V and Filipovic N, Accelerating Lattice-Boltzman method using Maxeler dataflow approach, The IPSI BgD Transactions on Internet Research, 34, 2013.
[8] Gnjatović M, Tasevski J, Mišković D, Savić S, Borovac B, Mikov A and Krasnik R, Pilot corpus of child-robot interaction in therapeutic settings, 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), IEEE, 253-258, September 2017.
[9] Tasevski J, Gnjatović M and Borovac B, Assessing the Children's Receptivity to the Robot MARKO, Acta Polytechnica Hungarica, 15(5):47-66, 2018.
[10] Kos A, Tomažič S, Salom J, Trifunovic N, Valero M and Milutinovic V, New benchmarking methodology and programming model for big data processing, International Journal of Distributed Sensor Networks, 11(8), 2015.
[11] Trifunovic N, Milutinovic V, Salom J and Kos A, Paradigm shift in big data supercomputing: dataflow vs. controlflow, Journal of Big Data, 2(1): 1-9, 2015.
[12] Montani S and Striani M, Artificial intelligence in clinical decision support: a focused literature survey, Yearbook of medical informatics, 28(1):120, 2019.
[13] Shickel B, Tighe PJ, Bihorac A and Rashidi P, Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis, IEEE journal of biomedical and health informatics, 22(5):1589–1604, 2018.
[14] Choi E, Bahadori MT, Searles E, Coffey C, and Sun J, Multi-layer Representation Learning for Medical Concepts, arXiv, 1–20, 2016.
[15] Van Der Maaten L and Hinton G, Visualizing Data using t-SNE, Journal of Machine Learning Research, 9:2579–2605, 2008.
[16] Che Z, Purushotham S, Khemani R, and Liu Y, Interpretable deep models for icu outcome prediction, AMIA Annual Symposium Proceedings, American Medical Informatics Association, p.371, 2016.
[17] Che Z, Purushotham S, Khemani R and Liu Y, Distilling knowledge from deep networks with applications to healthcare domain, arXiv preprint arXiv:1512.03542, 2015.
[18] Labrinidis A and Jagadish HV, Challenges and opportunities with big data, Proceedings of the VLDB Endowment, 5(12), 2012.
[19] Bojović M, Milutinović V, Bojić D and Korolija N, An Experimental Healthcare System: Essence and Challenges, Handbook of Research on Methodologies and Applications of Supercomputing, IGI Global, 301-308, 2021.