ARTIFICIAL INTELLIGENCE IN INTELLIGENT HEALTHCARE SYSTEMS-OPPORTUNITIES AND CHALLENGES

A. Petreska[1], B. Ristevski[2]

[1] Faculty of Information and Communication Technologies – Bitola
University "St. Kliment Ohridski"- Bitola, Republic of North Macedonia
e-mail: petreska.anita@uklo.edu.mk
[2] Faculty of Information and Communication Technologies – Bitola
University"St. Kliment Ohridski"- Bitola, Republic of North Macedonia
e-mail: blagoj.ristevski@uklo.edu.mk

Abstract:

Artificial intelligence (AI) has the potential to revolutionize healthcare by improving the accuracy, speed and efficiency of healthcare systems. Biomedical systems with the help of AI can improve the quality of healthcare by providing intelligent solutions that allow healthcare providers to make smart decisions, and generate new insights and discoveries. Biomedical systems use algorithms for Machine Learning (ML), Deep Learning (DL), and Neural Networks, and analyze large amounts of health data, including electronic health records, medical images, and patient histories.
AI algorithms can analyze patient data: genetic information, medical history, and lifestyle factors to develop personalized treatment plans tailored to each patient's unique needs. By analyzing patient data from a variety of sources, such as electronic health records, research articles, and real-time patient registries, AI algorithms can identify changes in a patient's condition and alert healthcare providers to potential problems before they become serious. AI algorithms analyze patient data, including clinical guidelines, treatment protocols, and patient-specific factors such as age, gender, and comorbidities, to recommend optimal treatment plans. AI plays a key role in remote patient monitoring and telemedicine, especially in underserved areas or during emergencies using virtual consultations. Implementing AI in healthcare requires collaboration between healthcare providers, technology companies and policy makers.

Key words: Machine Learning, Biomedical System, Medical Big Data, Algorithms

1. Introduction

In the modern era, the terms related to AI, machine learning and deep learning are used in fields where accurate prediction and data analysis are crucial [1]. In section 2, a reference is made to the efficient use of medical big data, - the characteristics - "10 Vs" of big data, Predictive, Preventative, Personalized, and Participatory Medicine (P4P), Electronic Health Record (EHRs), the most commonly used datasets that have the potential to transform healthcare delivery leading to improved patient care and outcomes, while ensuring appropriate data management, privacy and security measures comply with regulatory requirements [2-4]. Section 3 analyzes the most commonly used ML algorithms for medical data analysis [5-7]. Section 4 compares Spark and Hadoop, which are the leading open-source big data infrastructure software packages used to store and process large data sets [8]. The fifth section gives an overview of the open challenges and risks, which implies the implementation of a sophisticated health system in the context of big medical data [9]. Finally, in the sixth section, the concluding observations are presented [10].

Despite the numerous benefits of artificial intelligence in healthcare, there are also challenges that need to be addressed. Future work on AI for healthcare should focus on solving the challenges and expanding the scope of AI applications in healthcare.

2. Medical big data

Health data is characterized by a high degree of dimensional heterogeneity, untimeliness, and irregularity due to which the real value of these data is not fully utilized. AI enables improvements in the accuracy and speed of diagnosis, treatment, and advanced medical research that depend on the availability of large medical data sets. There are public datasets that are an essential starting point for scientists in the process of developing and training machine learning models. The Medical Information Mart for Intensive Care III (MIMIC-III is a publicly available database used to develop predictive models for mortality, sepsis and acute kidney injury, etc. The Surveillance, Epidemiology, and End Results (SEER) collects data on cancer incidence, treatment, and survival and is used to develop predictive models for cancer survival and to identify risk factors for developing cancer. The UK Biobank has collected genetic and lifestyle data to identify genetic variants associated with different diseases and develop predictive models for disease risk. The National Inpatient Sample is a database used to develop readmission prediction models. ImageNet is a large image database used to train models for deep learning and medical image analysis. Data science enables the development of P4M - Predictive, Preventative, Personalized, and Participatory Medicine [12]. HRs are owned by healthcare facilities where the patient is treated and there is almost no integration of this data when it comes to public and private health, and especially when it comes to different countries. The creation of a personal health record (PHR) owned and securely managed by the patient is a solution for providing health care and evidence-based medicine [12].

3. Analytical tools used for data mining

Big data is processed by clustering and scanning multiple cluster nodes in the network. Data mining is a multidisciplinary field that brings together database technology, statistics, ML and pattern recognition in the process of making real decisions [5] [7].

algorithm	**description**	**advantages**	**disadvantages**
Linear Regression	An easy way to figure out the connection between the inputs and a numerical output that changes in a steady way	- Explainable method - Interpretable results by its output coefficients - It takes less time	-Sensitive to outliers - Assumes linearity between inputs and output
Decision Tree	Decision trees create rules based on features to predict results They can be used for classifying or estimating values.	- Explainable and interpretable - Be able to deal with missing data	-Prone to overfitting - Sensitive to outliers
Random Forests	A bunch of decision trees working together is called a random forest.	- Reduces overfitting -This model has greater precision than other models	- Training complexity can be high - Not very interpretable
K-Means	Popular clustering method that looks at Euclidean distances to figure out the best way to separate data into K clusters	- Scales to large datasets - Simple to implement and interpret - Results in tight clusters	- Needing to figure out how many clusters should be expected beforehand. -It won't work for finding groups with non-traditional shapes.
Apriori algorithm	A method of finding the most commonly occurring itemset in a given set of data that uses information about the typical characteristics of frequent itemsets	- Results are intuitive and Interpretable - Trying to cover all the bases by finding all the rules based on the reliability and backing	- Needing to decide how many clusters or mix-ins you want before you start. - The covariance type needs to be defined for the mix of component

Table 2. Machine learningalgorithms

4. Software platforms for Big Data

Hadoop and Spark are software platforms, both developed by the Apache Software Foundation, are widely used open-source frameworks for big data architectures. Apache Hadoop is a free and open-source software framework used to store and process large data sets across distributed computing networks. It consists of several modules, including Hadoop Distributed File System (HDFS) for data storage and MapReduce for data processing. Apache Spark is an open-source distributed computing system for big data processing and analytics. It provides a unified API for distributed data processing that supports various programming languages such as Scala, Python, Java, and R. Spark includes various libraries and tools such as Spark SQL for processing structured data [8].

Apache Spark	Apache Hadoop
Easy to program and does not require any abstractions	Difficult to program and requires abstractions
Developers can perform streaming, processing and machine learning, all in the same cluster	It is used to generate reports
It has a built-in interactive mode.	There is no built-in interactive mode, except for tools like Pig and Hive.
Runs 10 to 100 times faster than Hadoop MapReduce.	Hadoop MapReduce does not use Hadoop cluster memory to its maximum
Developers can modify data in real-time via Spark streaming	Allows to process a group of stored data.

Table 1. Apache Spark vs Apache Hadoop.

5. Challenges and risks

Artificial intelligence is transforming the healthcare industry by enabling faster and more accurate diagnosis, personalized treatment and efficient allocation of resources. However, there are still several challenges that need to be addressed for AI to be fully integrated into intelligent healthcare systems [9]. Data integration enables the integrity of heterogeneous data sources, such as electronic health records (EHRs), medical imaging, genomics, and wearable devices. To overcome this challenge, several approaches have been developed. Data normalization - converting data into a standardized Health Level Seven International (HL7) format. Data mining is the process of discovering patterns and relationships in large data sets used to extract meaningful insights from EHRs and medical images; Federated learning is a machine learning technique that enables local training and analysis of data on different devices while preserving data privacy, which is particularly useful for analyzing medical data subject to privacy regulations. Medical data is subject to strict regulations, such as HIPAA (Health Insurance Portability and Accountability Act) in the US and GDPR and EHDS in Europe. Another challenge is the quality of medical data which is often incomplete, inaccurate or inconsistent, which can affect the performance of AI algorithms. To meet this challenge, several approaches have been developed: data preprocessing, feature selection and data augmentation. Another challenge in AI in healthcare is the lack of explainability of AI algorithms. To address this challenge, several approaches have been developed: model interpretation, rule-based systems and hybrid approaches.

6. Conclusion and future work

Big data analysis mainly focuses on the perspectives of machine learning for personalized medicine, genomic data models and the application of data mining algorithms. It is encouraging that deep learning methodologies have improved predictive models, however, the interpretability of such models remains a challenge in the future. AI cannot multitask, and a full replacement for a physician has yet to be developed. Natural language processing (NLP) algorithms can be used to identify key phrases, medical terms and other relevant information in clinical notes, enabling the analysis of large amounts of data. In healthcare, computer vision is used to analyze medical images, such as X-rays, CT scans and MRIs. Computer vision algorithms can be trained to identify abnormalities in medical images that may indicate disease. Healthcare robotics can be used to automate routine tasks, such as drug delivery and patient monitoring. It is necessary to choose an appropriate big data platform that will have all the required libraries including machine learning libraries. The rapid development of healthcare applications raises privacy concerns, so the blockchain technology could be a good solution. To increase the formality and standardization of data mining methods, a new programming language may need to be developed specifically for this purpose, as well as new methods capable of addressing unstructured data such as graphics, audio, and handwritten text. Due to the continuous growth of healthcare costs, it is necessary to strike a balance between costs and quality of services [10].

References

[1] Aggarwal, Karan, et al. "Has the future started? The current growth of artificial intelligence, machine learning, and deep learning." Iraqi Journal for Computer Science and Mathematics 3.1 (2022): 115-123.

[2] Savoska, Snezana, Blagoj Ristevski, and Vladimir Trajkovik. "Personal Health Record Data-Driven Integration of Heterogeneous Data." Data-Driven Approach for Bio-medical and Healthcare. Singapore: Springer Nature Singapore, 2022. 1-21.

[3] Nti, Isaac Kofi, et al. "A mini-review of machine learning in big data analytics: Applications, challenges, and prospects." Big Data Mining and Analytics 5.2 (2022).

[4] Savoska, Snezana, Blagoj Ristevski, and Vladimir Trajkovik. "Personal Health Record Data-Driven Integration of Heterogeneous Data." Data-Driven Approach for Bio-medical and Healthcare. Singapore: Springer Nature Singapore, 2022. 1-21.

[5] Ristevski, Blagoj, and Ming Chen. "Big data analytics in medicine and healthcare." Journal of integrative bioinformatics 15.3 (2018).

[6] Kumar, Sunil, and Maninder Singh. "Big data analytics for healthcare industry: impact, applications, and tools." Big data mining and analytics 2.1 (2018): 48-57.

[7] Hassanien, Aboul-Ella, Kuo-Chi Chang, and Tang Mincong, eds. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2021. Vol. 1339. Springer Nature, (2021).

[8] Aziz, Khadija, Dounia Zaidouni, and Mostafa Bellafkih. "Real-time data analysis using Spark and Hadoop." 2018 4th international conference on optimization and applications (ICOA). IEEE, (2018).

[9] Wani, Shahid Ud Din, "Utilization of artificial intelligence in disease prevention: Diagnosis, treatment, and implications for the healthcare workforce, MDPI, (2022).

[10] Blaizot, Aymeric, et al. "Using artificial intelligence methods for systematic review in health sciences: A systematic review." Research Synthesis Methods 13.3 (2022).

[11] Esther, C., et al. "State-of-the-art recent technologies, background, terminologies, and challenges: a preliminary survey on big data." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 13.2 (2022): 74-77.

[12] Stipelman, Carole H., et al. "Electronic Health Record-Integrated Clinical Decision Support for Clinicians Serving Populations Facing Health Care Disparities: Literature Review." Yearbook of Medical Informatics 31.01 (2022): 184-198.

[13] Hathaliya JJ, Tanwar S, An exhaustive survey on security and privacy issues in Healthcare 4.0. Comput Commun 153:311–335, (2020).