Big Data-Based Optimized Deep Learning Model for Improving Performance of Electronics Health Care Data

Rajeev Kumar Bhaskar [1], Balasubramaniam Kumaraswamy[2]

[1]*Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.*
[2]*Research Scholar, Department of CS & IT, Kalinga University, Raipur, India.*

KEYWORDS

Health care, Big Data, Deep learning.

ABSTRACT

Bigdata analytics is a new area of supervised analytics that healthcare analytics has moved into. Healthcare data is transmissive, incremental, and substantial, and it has pre-set thresholds for classifying patient conditions and diseases. The necessary understanding of disease occurrences is applied to the analysis of this data. The research project introduces healthcare data analytics with a novel framework that leverages unstructured data for the classification of healthy and unhealthy samples in order to understand the nature of healthy and unhealthy labelled classes. This is motivated by the fact that healthcare data is informative and voluminous in nature. Additionally, the unknown samples' health status is predicted using the supervised knowledge. Each model presented in this work is supported by accuracy and a computational complexity model.

1. Introduction

Due to the vast amount of medical records that are included in patient health records, big data has become a crucial component of healthcare. Medical records are being digitalised and kept up to date as Electronic Medical Records (EMR) in the current medical landscape. Due to the volume of these records, research on the identification of helpful patterns about side effects, drug effectiveness, and associated data can be conducted [1]. These reports can also provide an overview of the course of treatment, including any procedures performed and any unfavourable outcomes. Drug side effects can occasionally be deadly and dangerous. The suggested approach is identifying these impacts, and a medicine process study is being conducted [2]. The primary goal of data mining is to extract information from a data set and transform it into a usable format for later use [6]. Presumably, it is the Knowledge Discovery in Databases (KDD) process's intermediate step, which focusses on identifying more dependable patterns and methods that aid in data type prediction.

A few other phases in KDD include selecting, cleaning, and preparing data as well as combining pertinent data and evaluating the findings of the mining process. In diverse disciplines of research such as statistics, machine learning, mathematics, artificial intelligence, and pattern recognition using specific models, different DM strategies are developed and implemented [11]. Big data from the medical field differs from big data from other fields in a number of key ways. Big data in medicine is often difficult to access, and most medical researchers are reluctant to employ open data science due to concerns about data exploitation by third parties and a lack of incentives for sharing data [9]. Because the extraction procedure makes raw data simpler, medical big data are frequently gathered using protocols (i.e., fixed formats) and are somewhat structured. The fact that medicine is performed in a situation where patient safety is paramount and explanations must accompany decision-making is another crucial aspect. [4]. The rest of the paper is organized as follows: Section 2 provides the classification scheme for the survey; Section 3 provides an overview of proposed architecture. Section 4 provides a summary and comparison of the results of the various papers discussed in this taxonomy. Finally, Section 5 concludes the paper.

2. Related Works

Big data in medicine can be expensive because it requires staff to use pricey equipment and may cause discomfort for the patients. When compared to data from other fields, medical big data are modest and can come from nonreproducible sources. Several causes of uncertainty, including measurement errors, missing data, or mistakes in coding the information hidden in textual reports, might further impact medical big data. As a result, both data analysis and result interpretation may heavily rely on domain expertise [5]. The various patient characteristics, which occasionally may need to be weighted; the time structure, which could be an extra dimension; and the treatment information, time point of treatment decision and change (i.e., time-dependent confounding) are additional unique features of medical big data in analytical aspects. [12]. A secure and private blockchain-based EHR sharing system is suggested by Yong Wang et al. In order to ascertain the effect of EHR on population health, Clemens Scott Kruse et al. [4] conducted a review in which they identified and analysed the factors that promote and hinder the use of EHRs. Amidst the vast array of electronic health records (EHRs) that have created immense opportunities for enhancing healthcare, two essential textual modes are clinical data (structured data) and clinical narratives (unstructured data). The majority of EHRs now in use concentrate on just one modality or superficially merge data from several sources, neglecting the innate connections between them. To overcome these challenges, [5] introduced a Medical Multimodal Pre-Trained Language Model, or MedM-PLM, to learn more accurate EHR descriptions over structured and unstructured data. It is clear from the poll that health care technology has advanced from EHR/EMR in its infancy to HIT. The vast quantity of data made accessible by medical and health care equipment has encouraged technological researchers to create automated gadgets that combine smart and medical technologies. Health equipment with reliable brands that can identify, recognise, and capture photographs are available in the corporate hospital and health care markets for usage by physicians. [6]. It is often acknowledged that the clinical development of pharmaceuticals is an intricate procedure that takes a very long time to complete.

The first phase in the digital technology transformation process is to leverage health care data, which necessitates data flow amongst various stakeholders in the health care community [3]. Healthcare data is a collection of both public and private information that is utilised by a variety of organisations, including insurance companies, hospitals, and physicians. It comprises administrative enrolments, health surveys, and medical records. Poor data quality has a detrimental effect on validity, poor patient care, and study findings can replicate misleading claims about a variety of outcomes. In the field of healthcare analytics, the quality of the input data is crucial to achieving high accuracy in any diagnosis model. Healthcare is an example of Medical-Big Data [M-BD], which presents heterogeneity (diversity) with volume and velocity—a crucial component of big data. This new functionality necessitates an effective framework that provides correct data replacement with proper modelling of the database. With the integration of big data into healthcare, a unique approach has been proposed in this study to discover and correct cross- or missing-data in all database positions. This allows for the ordering of data points and the identification of actionable insights. [7].

3. Methodologies

Current data analysis and administration are heavily reliant on new technology and tools to support the big data characteristics. Despite this, there has been a rise in new research in the era of digital health care due to the quick growth and widespread accessibility of healthcare data, with a shift in medical practice towards prevention rather than treatment. Hospitals and medical families generate a multitude of diverse clinical data through medical imaging, sensors, and patient records. This can be viewed as a rich source of big data that can be leveraged to enhance medical systems. Health Information Technology (HIT), which organises data without errors, noise, or missing data, has added a new dimension to health care data analytics. Healthcare data is vast and contains a great deal of characteristics and samples that are time, location, and event-recorded. Reducing the dimension of the data by extracting relevant features has a significant effect on the computation and space complexity of any intelligent algorithm used later in analytics. This study work has used the dimensionality reduction technique with feature selection and extraction in order to reduce complexity [8]. The original feature space is shrunk to a lower dimensional version using feature extraction. Algorithms like Independent Component Analysis (ICA) are frequently used for feature extraction. However, if feature vectors are dispersed along a nonlinear manifold in a high-dimensional space, ICA, being linear projection methods, could lead to errors in classification. The process of choosing the best set of features from the original features while keeping enough information is known as feature selection.

The developed feature selection techniques are called Particle Swarm Optimisation (PSO). GAs are a well-liked option among them. Nevertheless, in certain scenarios of multifunctional optimisation, GAs were unable to maintain several global or regional optima. Several attempts have been made to modify the fitness competence rule and incorporate scaling fitness in order to improve GAs' capacity to reach multiple peak solutions. The DCNN method can be repeatedly executed by the computer once the input attributes are mapped to the intended output category field. Using deep learning techniques, the user can input the values of multiple attributes, from which his risk factor will be computed. This is a nonlinear combination of notable features where each component is coupled with very other feature. The last layer is a SoftMax layer that is stacked at the end for data classification, after which comes the FC layer output and stacked or deep multiple layers. The suggested GA-based DCNN works effectively to reduce false positives and raises the classifier's accuracy. [10].

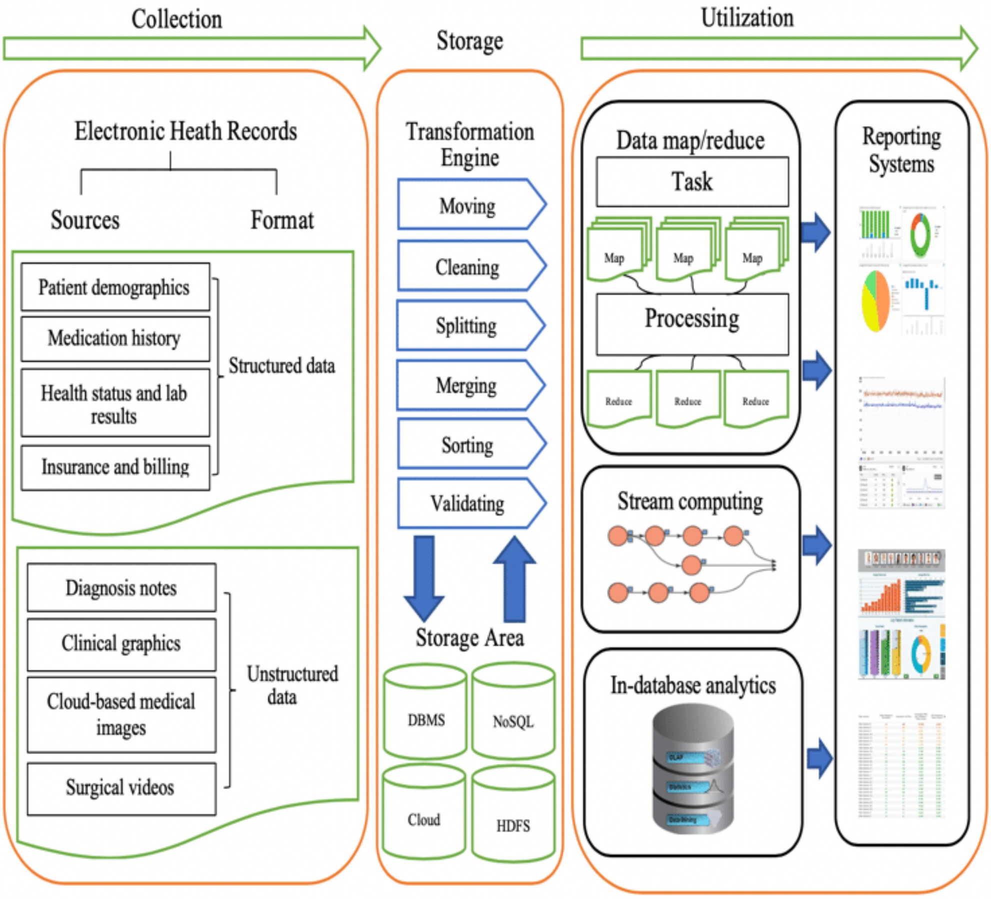

Figure 1. Framework of Proposed Method

4. Results and discussion

A comparison was made between the unanalysed and analysed attributes in the digging model in terms of processing time and evaluation accuracy. The outcomes shown in Table 1. Figure 3 shows that the decision tree method's data mining accuracy rate is almost 96%, which is either close to or higher than the impact of a clinical diagnosis. This clearly demonstrates the efficacy of our suggested algorithm.

Table 1. Comparison table mining model results

Processing Quantity	**Without optimization**		**With optimization**	
	Processing Time (ms)	**Accuracy of evaluation (%)**	**Processing Time (ms)**	**Accuracy of evaluation (%)**
50	754	75	878	80
100	875	79	936	85

150	926	80	969	89
200	957	85	970	90
250	976	89	1084	93
300	1084	90	1989	96

Out of the 76 features/attributes in the original dataset, we used the dimensionality reduction technique to select 14 for further analysis. The data set is partitioned into 70% training data and 30% testing data in order to assess the performance of the suggested algorithms.

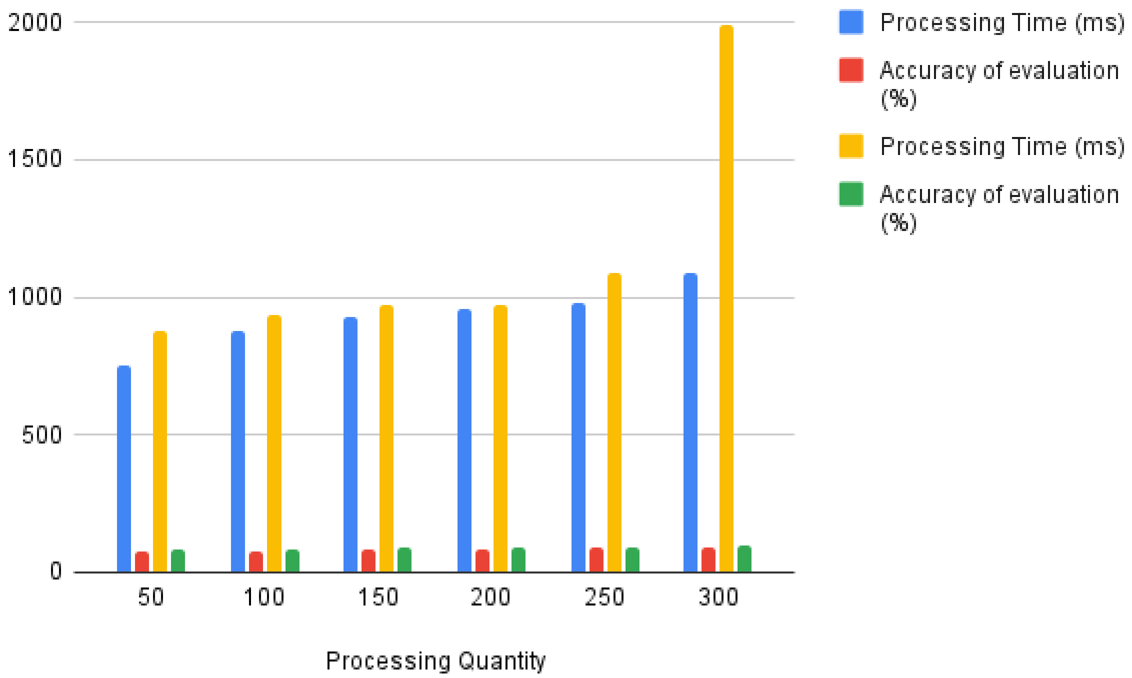

Figure 2. The Accuracy of proposed model

5. Conclusion and Future Scope

Health care is the systematic provision of medical treatment to people or communities with the goal of ensuring optimal health. Digital healthcare technologies, such as 3D printing, robotic machines, nanotechnology, virtual reality, artificial intelligence, and virtual reality, can be used to provide high-quality healthcare services. They will significantly affect health care in the future. Innovation in technology is needed in hospitals and clinics to minimise the usage of human resources. Many technology researchers have been motivated by this need to develop sophisticated algorithms that identify, categorise, and forecast occurrences in health-care data. In this work, the GA-based DCNN is utilised. The proposed model has great accuracy. The low computing complexity of the DCNN algorithm gives it an advantage over today's most popular deep architectures.

References

[1] Pramanik, Pijush Kanti Dutta, Saurabh Pal, and Moutan Mukhopadhyay. "Healthcare big data: A comprehensive overview." *Research anthology on big data analytics, architectures, and applications* (2022): 119-147.

[2] Manogaran, Gunasekaran, Daphne Lopez, Chandu Thota, Kaja M. Abbas, Saumyadipta Pyne, and Revathi Sundarasekar. "Big data analytics in healthcare Internet of Things." Innovative healthcare systems for the 21st century (2017): 263-284.

[3] S. Neelima, Manoj Govindaraj, Dr.K. Subramani, Ahmed ALkhayyat, & Dr. Chippy Mohan. (2024). Factors Influencing Data Utilization and Performance of Health Management Information Systems: A Case Study. Indian Journal of Information Sources and Services, 14(2), 146–152. https://doi.org/10.51983/ijiss-2024.14.2.21

[4] Karatas, Mumtaz, Levent Eriskin, Muhammet Deveci, Dragan Pamucar, and Harish Garg. "Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives." Expert Systems with Applications 200 (2022): 116912.
[5] Firouzi, Farshad, Amir M. Rahmani, Kunal Mankodiya, Mustafa Badaroglu, Geoff V. Merrett, P. Wong, and Bahar Farahani. "Internet-of-Things and big data for smarter healthcare: From device to architecture, applications and analytics." Future Generation Computer Systems 78 (2018): 583-586.
[6] Majeed, A.S. Eco-friendly design of flow injection system for the determination of bismarck brown R dye (2018) International Journal of Pharmaceutical Research, 10 (3), pp. 399-408.
[7] Madhavi, M., Sasirooba, T., & Kumar, G. K. (2023). Hiding Sensitive Medical Data Using Simple and Pre-Large Rain Optimization Algorithm through Data Removal for E-Health System. Journal of Internet Services and Information Security, 13, 177-192.
[8] Hong, Liang, Mengqi Luo, Ruixue Wang, Peixin Lu, Wei Lu, and Long Lu. "Big data in health care: Applications and challenges." Data and information management 2, no. 3 (2018): 175-197.
[9] Karatas, Mumtaz, Levent Eriskin, Muhammet Deveci, Dragan Pamucar, and Harish Garg. "Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives." Expert Systems with Applications 200 (2022): 116912.
[10] Mohamed, K.N.R., Nijaguna, G.S., Pushpa, Dayanand, L.N., Naga, R.M., & Zameer, AA. (2024). A Comprehensive Approach to a Hybrid Blockchain Framework for Multimedia Data Processing and Analysis in IoT-Healthcare. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 15(2), 94-108. https://doi.org/10.58346/JOWUA.2024.I2.007
[11] Wang, Lidong, and Cheryl Ann Alexander. "Big data analytics in medical engineering and healthcare: methods, advances and challenges." Journal of medical engineering & technology 44, no. 6 (2020): 267-283.
[12] El Khatib, Mounir, Samer Hamidi, Ishaq Al Ameeri, Hamad Al Zaabi, and Rehab Al Marqab. "Digital disruption and big data in healthcare-opportunities and challenges." ClinicoEconomics and Outcomes Research (2022): 563-574.
[13] Dimitrov, Dimiter V. "Medical internet of things and big data in healthcare." Healthcare informatics research 22, no. 3 (2016): 156-163.