Healthcare Data Analytics and Privacy Preservation by DCNN Algorithm

Dr. Uruj Jaleel[1], R Lalmawipuii[2]

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

KEYWORDS
Healthcare, Data Analytics, DDS, deep Learning, Classification

ABSTRACT
Data has become an integral part of the digital world with the advancement in computing technologies. The collection of data is very crucial with regards to data analytics. Every industry makes use of data analytics ranging from financial to other commercial applications but it becomes even more important in healthcare domain for the analysis of healthcare data. The present research work is mainly focused on classification/prediction problems of healthcare data based on deep learning (supervised) approaches using data mining techniques. There is a need to design an intelligent model (based on deep learning) which can classify the amount of data that is stored in our databases. Human data analytical capability rate is much smaller when compared to the amount of data that is stored. This (classification) becomes even more critical when it comes to healthcare data as it can help to detect, diagnose and treat the patients based on these classified data. The main goal of the thesis is to develop a deep learning-based model for classification tasks and the introduced DDS can be used in healthcare domain to improve the diagnostic speed, accuracy and reliability.

1. Introduction

Medical disease diagnosis systems, or MDDSs, are presently being investigated as a way to assist clinicians, who are the front-line healthcare providers, in making decisions about illness diagnosis and treatment. This could improve patient outcomes and save costs. Several clinical factors are taken into consideration when making these healthcare decisions [1]. Physicians typically treat patients based on their clinical expertise and personal experience. Because medical professionals have varying degrees of experience, it is possible for them to diagnose patients incorrectly at times and for manual treatment to take longer. Additionally, as the population grows, there is a daily increase in need for medical professionals. Computer-based MDDSs are therefore required in order to enable physicians to make prompt and more knowledgeable healthcare decisions [2]. In order to do this, these computerised systems are being employed widely as a promising tool to help doctors diagnose patients, improving the standard of care and eventually cutting expenses [19]. MDDSs are becoming an essential component of the healthcare system [14].

But there are a number of problems and real-world limitations with health-related data. Most notably, a problem with data imbalance in the medical datasets comes when a certain class of disorders occurs in very few patients [3]. Furthermore, it can be difficult to continuously find a significant number of patients to represent some diseases due to their low prevalence [4]. Third, there are contradicting difficulties since persons with the same medical disease do not all appear the same clinically [12]. These are the principal problems encountered in the process of designing an effective computerised MDDS. Clinical data pertaining to illnesses are generally utilised extensively. It is connected to each and every person as well. The human species is always struggling to discover cures for the various illnesses and symptoms that appear every day. Many diseases have established diagnoses, symptoms, and potential treatments; yet, these details are scarce and not available to everyone worldwide . In addition, there are a number of other drawbacks, including the model's disease specificity, incomprehensibility, lack of decision power, and inadequate dimensionality of clinical data [20]. These are the reasons that creating a precise, accurate, and dependable predictive model—which helps physicians and clinical specialists diagnose and treat patients more effectively—is the hardest undertaking there is. Creating an efficient MDDS will help healthcare practitioners manage patient data more easily, which is the aim of the current research [15].

In this case, the introduction is examined in section 1 of the article while the pertinent literature is examined in section 2. Section 3 explains the plan for the work, Section 4 shows the results of the work, and Section 5 wraps up the project.

2. Literature Review

This section covers machine learning techniques for managing healthcare data effectively, supporting decision-making and providing insights that may be put into practice. All health-related digital records are generally referred to as healthcare data [6]. It could include comprehensive data regarding the medical histories of the patients, prescription notes from the doctors, clinical reports, etc [13]. Healthcare big data is the product of all these large-scale, highly dimensional, and diverse data sets [16]. These data come from a variety of internal and external sources, including social media, biometric, healthcare, and picture data. An additional significant challenge facing today's Healthcare Information Systems (HISs) is the daily exponential growth in healthcare data [18]. In addition to the enormous volume of healthcare data, this transition is marked by a sharp increase in the velocity and diversity of data generation [5]. The term "healthcare big data" refers to the notable similarities between the nature and properties of big data and healthcare data [11]. Clinical diagnostic systems are created in the medical industry to help doctors make decisions about patient care based on health records [8].

By including the patient's medical history in the planning of safe and appropriate care, electronic health records, or EHRs, are said to empower clinicians ([9]. A thorough investigation on the use of data mining in healthcare and medicine was conducted by In. [10], and it focused on six medical tasks: prognosis, monitoring, therapy, screening, diagnosis, and management [7]. A methodology for classifying brain tumours and estimating their level of malignancy, utilising multi-parametric imaging profile data from CT, conventional MRI, and advanced MRI scans, was shown in [17]. Despite the large number of research studies conducted in this field, much work still needs to be done. In response to the shortcomings of the current models, this work offers a stable, optimal feature-selection technique that primarily targets various-sized medical datasets. Stability is specifically defined as how sensitive the selected feature set is to changes in the training dataset that is provided. more precisely, the degree to which the dataset's little changes affect the model. Such a concept is undoubtedly significant, particularly for medical datasets.

3. Methodology

Three steps make up the suggested model: STAGE-I, STAGE-II, and STAGE-III. The introduced model's first phase is dedicated to gathering data. The processing of data comes next, and building the illness diagnosis system model comes last.

These are units within the network of hospitals that are managed by a single or several healthcare professionals, including nurses, general practitioners, and other staff members. Their goal is to offer general healthcare to local residents. These primary centres' principal objective is to treat individuals with brain tumours while also offering general healthcare services. This becomes even more crucial in India's rural areas, where a higher death rate results from individuals trying to avoid contacting these healthcare facilities and their personnel out of ignorance and lack of information. The photograph was first acquired from the Community Health Centre. Next, the brain dataset that was gathered from the Community Health Centre is pre-processed using machine learning-based data discretization.

Since classifiers prefer to handle discrete data over continuous values when learning, data discretization is essential to the machine learning process. This is further corroborated by the finding that using data discretization improves the quality of discovered information. It is also predicated on the idea of information theory. Classification is the most well-known use of deep learning algorithms. The current study also uses Deep Convolutional Neural Network algorithms to forecast disease diagnosis systems. The augmentation is performed using this deep CNN.

Fully linked layers, which establish connections between each neuron in a layer and every other layer's neuron, are a common feature of neural networks. This setup allows fast handling of correlations between any points in the training vectors, regardless of their proximity to one another. However, when crucial information is not restricted to a local context, Convolutional Neural Networks (CNNs) perform less effectively because they are built to cope with local structures. The fact that CNNs' convolutional layers have fewer parameters than fully connected layers gives them an advantage in training [11–13].

Given that local features like edges and gradients are commonly visible in images, CNNs are highly helpful for image analysis. CNNs have demonstrated efficacy in a range of problem domains outside images. Any situation where the local structures are emphasised in the data representation is a suitable fit for CNN. This feature applies to text and speech analysis in addition to images. In addition to convolutional layers, CNN architectures often contain pooling layers. Pooling layers, among other things, enable down sampling from a specific layer, enhancing the efficiency of subsequent computations. The obtained features are converted into vector format for the classifier using a feature selection technique.

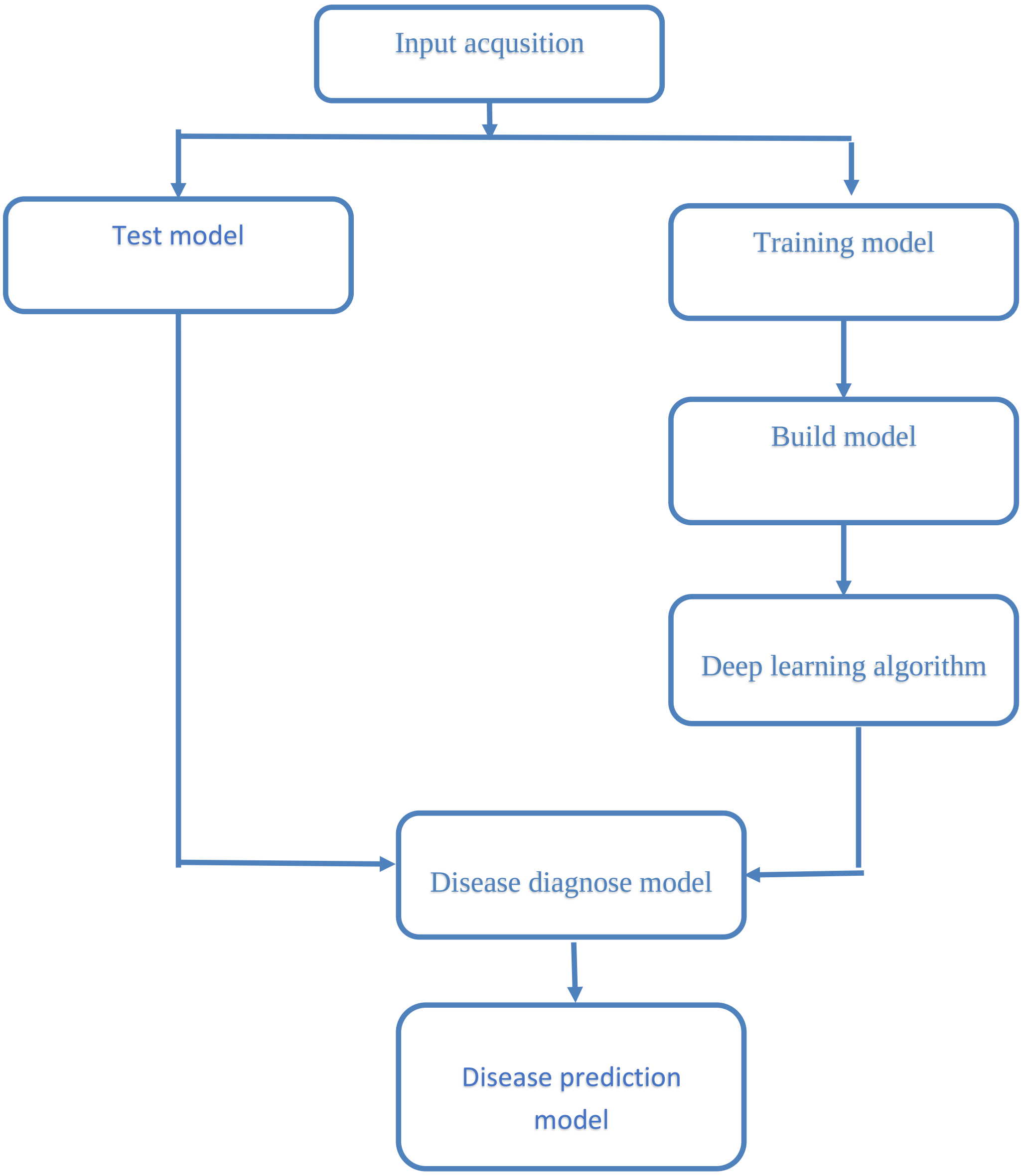

Figure 1: Schematic Diagram of proposed framework

The features dictate the separation between the number of clusters and the anchoring of deep CNNs. This disagreement is given an activation function by the evaluation, which is also referred to as inferences in the deep learning model. The activation levels are then divided by the strengths of the output nodes to determine the network outcomes. In particular, this evaluation technique is:

$$r^2 = ||y^{\mu} - C_k||^2_{R_k} = (y - C_k)^T R_k (y - C_k) \qquad (1)$$

$$Z_i = F_i(y) = \sum_{k=1}^{C} W_{ki} h\left(||y^{\mu} - C_k||^2_{R_k}\right) + W_{0i} \qquad (2)$$

Where h is the activation function, W_{0i} are the biases, W_{ki} is the output layer's weight, T is the matrix transportation, and r stands for distance. R_k is the positive definite covariance matrix. The numbers of samples, cluster centres, and classes are listed in the aforementioned equations by μ, k, and i.

4. Results and discussion

The input dataset was divided into training and test datasets in order to facilitate model fitting. Thirty percent of the input dataset is designated for testing and seventy percent is used for training. The training set is used in the learning phase to find and select the optimal model, whereas the test dataset is used in the prediction phase. The project makes use of a brain dataset containing observations from 1500 patients. The experiment is conducted using the WEKA toolkit or Python and Scikit-Learn package.

Based on specificity, accuracy, and sensitivity, the effectiveness of several classifiers utilised in this work, including RNN, LSTM, DBN, ResNet, and Deep CNN, is assessed. These metrics can be used to compare the three suggested classifiers with the proposed model performance. Tables 1 and 2 display the various classifiers' performance metrics for the original and balanced datasets, respectively.

Table 1. Performance Metrics of the Classifiers-Original Dataset

Models	Accuracy	F1-Score	Precision	Recall
RNN	75	0.87	87	89.85
LSTM	88	0.85	87.50	90.25
DBN	86	0.87	88.15	81.05
RseNet	91	0.85	87.50	90.25
DCNN	93.5	0.92	92.15	91.05

Table 2. Performance Metrics of the Classifiers- Balanced Dataset

Models	Accuracy	F1-Score	Precision (%)	Recall (%)
RNN	87.5	0.90	87	85.85
LSTM	89.2	0.92	88	82.25
DBN	86.5	0.87	89	81.05
RseNet	92	0.91	88.9	88.25
DCNN	98.9	0.97	98.15	97.05

When the accuracy rates of the three classifiers listed above are compared, it is clear that DCNN is the best learner, since it can attain a 98.9% accuracy rate. The intended model's primary goal is to assist medical practitioners in making decisions while providing patient care. It is accomplished by determining the likelihood of problems during diagnosis, cutting down on the expense of diagnosis, and ultimately lowering the danger of death.

Further for the identification of MITM attacks, DCNN algorithms provided better accuracy values of 98.5% as shown in figure 2.

Table 3: Performance metrics of attack detection

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	Computational Time (sec)	Memory Utilization (%)
RNN	82.15	88.88	68.25	0.91	576.38	75.21
LSTM	95	97	89.85	0.97	193.39	45.25
DBN	96.50	97.50	90.25	0.985	101.75	40.25
DCNN	98.7	98.15	81.05	0.99	99.92	41.90

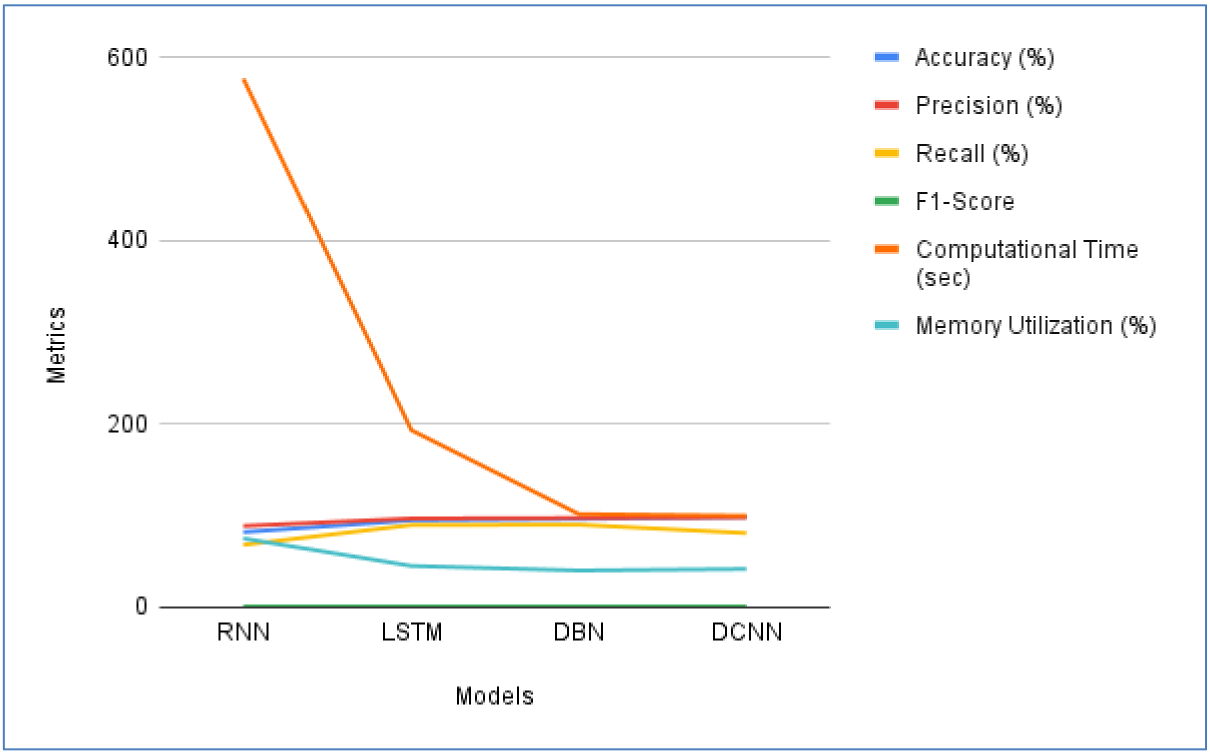

Figure 2: Performance metrics of attack detection

The effectiveness of our suggested models was assessed using accuracy metrics. Our suggested model's accuracy was really satisfactory. We contrasted our findings with those attained using alternative techniques, as seen in Figure 2.

5. Conclusion and future scope

A deep learning-based model for brain tumour prediction was created in the current study. The key features in the brain dataset can be found using the recently presented model. Deep learning classifiers, specifically DCNN, are used for this purpose in order to distinguish between benign and malignant brain tumours. When the accuracy rates of the three classifiers listed above are compared, it is clear that DCNN is the best learner, since it can attain a 98.9% accuracy rate. The intended model's primary goal is to assist medical practitioners in making decisions while providing maternity care. It is accomplished by determining the brain tumour risk. The deep learning-based model that was built is used in the healthcare industry to increase the reliability, accuracy, and speed of diagnosis. By creating MDDS for healthcare big data using big data analytical tools and approaches, future research can be investigated. The DCNN algorithm, which detects MITM attacks with 98.7% accuracy, is likewise concerned about security.

Reference

[1] Bi, Hongliang, Jiajia Liu, and Nei Kato. "Deep learning-based privacy preservation and data analytics for IoT enabled healthcare." *IEEE Transactions on Industrial Informatics* 18, no. 7 (2021): 4798-4807.

[2] Alguliyev, Rasim M., Ramiz M. Aliguliyev, and Fargana J. Abdullayeva. "Privacy-preserving deep learning algorithm for big personal data analysis." *Journal of Industrial Information Integration* 15 (2019): 1-14.

[3] Chatterjee, P., Siddiqui, S., Granata, G., Dey, P., & Abdul Kareem, R. S. (2024). Performance Analysis of Five U-Nets on Cervical Cancer Datasets. Indian Journal of Information Sources and Services, 14(1), 17–28.

[4] Liu, Yang, Zhuo Ma, Ximeng Liu, Siqi Ma, and Kui Ren. "Privacy-preserving object detection for medical images with faster R-CNN." *IEEE Transactions on Information Forensics and Security* 17 (2019): 69-84.

[5] Bae, D., & Ha, J. (2021). Performance Metric for Differential Deep Learning Analysis. Journal of Internet Services and

Information Security, 11(2), 22-33.

[6] Ali, Mansoor, Faisal Naeem, Muhammad Tariq, and Georges Kaddoum. "Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey." *IEEE journal of biomedical and health informatics* 27, no. 2 (2022): 778-789.

[7] Malathi, K., Shruthi, S.N., Madhumitha, N., Sreelakshmi, S., Sathya, U., & Sangeetha, P.M. (2024). Medical Data Integration and Interoperability through Remote Monitoring of Healthcare Devices. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications(JoWUA),15(2), 60-72. https://doi.org/10.58346/JOWUA.2024.I2.005

[8] Thilagam, K., A. Beno, M. Vanitha Lakshmi, C. Bazil Wilfred, Santhi M. George, M. Karthikeyan, Vijayakumar Peroumal, C. Ramesh, and Prabakaran Karunakaran. "Secure IoT Healthcare Architecture with Deep Learning-Based Access Control System." *Journal of Nanomaterials* 2022, no. 1 (2022): 2638613.

[9] Huang, Qi-Xian, Wai Leong Yap, Min-Yi Chiu, and Hung-Min Sun. "Privacy-preserving deep learning with learnable image encryption on medical images." *IEEE Access* 10 (2022): 66345-66355.

[10] Almaiah, Mohammed Amin, Fahima Hajjej, Aitizaz Ali, Muhammad Fermi Pasha, and Omar Almomani. "A novel hybrid trustworthy decentralized authentication and data preservation model for digital healthcare IoT based CPS." *Sensors* 22, no. 4 (2022): 1448.

[11] Arora, G. (2024). Desing of VLSI Architecture for a flexible testbed of Artificial Neural Network for training and testing on FPGA. Journal of VLSI Circuits and Systems, 6(1), 30-35.

[12] Moqurrab, Syed Atif, Noshina Tariq, Adeel Anjum, Alia Asheralieva, Saif UR Malik, Hassan Malik, Haris Pervaiz, and Sukhpal Singh Gill. "A deep learning-based privacy-preserving model for smart healthcare in Internet of medical things using fog computing." *Wireless Personal Communications* 126, no. 3 (2022): 2379-2401.

[13] Jelena, T., & Srđan, K. (2023). Smart Mining: Joint Model for Parametrization of Coal Excavation Process Based on Artificial Neural Networks. Archives for Technical Sciences, 2(29), 11-22.

[14] Kathamuthu, Nirmala Devi, Annadurai Chinnamuthu, Nelson Iruthayanathan, Manikandan Ramachandran, and Amir H. Gandomi. "Deep Q-learning-based neural network with privacy preservation method for secure data transmission in internet of things (IoT) healthcare application." *Electronics* 11, no. 1 (2022): 157.

[15] Saravanan T., et.al A protected multi authorizing reprogramming protocol for wireless sensor networks, Middle - East Journal of Scientific Research, V-20, I-12, PP:2554-2558, 2014.

[16] Ge, Chunpeng, Changchun Yin, Zhe Liu, Liming Fang, Juncen Zhu, and Huading Ling. "A privacy preserve big data analysis system for wearable wireless sensor network." *Computers & Security* 96 (2020): 101887.

[17] Almaiah, Mohammed Amin, Aitizaz Ali, Fahima Hajjej, Muhammad Fermi Pasha, and Manal Abdullah Alohali. "A lightweight hybrid deep learning privacy preserving model for FC-based industrial internet of medical things." *Sensors* 22, no. 6 (2022): 2112.

[18] Tabassum, Aliya, Aiman Erbad, Amr Mohamed, and Mohsen Guizani. "Privacy-preserving distributed IDS using incremental learning for IoT health systems." *IEEE Access* 9 (2021): 14271-14283.

[19] Kutlu, Y., & Camgözlü, Y. (2021). Detection of coronavirus disease (COVID-19) from X-ray images using deep convolutional neural networks. Natural and Engineering Sciences, 6(1), 60-74.

[20] Alzubi, Jafar A., Omar A. Alzubi, Ashish Singh, and Manikandan Ramachandran. "Cloud-IIoT-based electronic health record privacy-preserving by CNN and blockchain-enabled federated learning." IEEE Transactions on Industrial Informatics 19, no. 1 (2022): 1080-1087.