MULTI-DISEASE DETECTION WITH NEURAL NETWORK FUSION

Manoj Kumar Singh, Simran Arora, Shobhit Saini, Faraz Jafri, Wazahat Ali

[1]Department of CSE, MIT Moradabad, Moradabad, INDIA
manojaswal1982@gmail.com
simran.arora1271@gmail.com
shobhitsaini029@gmail.com
Official8ali@gmail.com
Jafrifaraz@gmail.com

ABSTRACT

In this study, we address the challenge of analyzing imbalanced medical data for accurate disease prediction. Specifically, we focus on developing a supervised model that utilizes machine learning and deep learning techniques to detect multiple diseases. In our approach, we focus on chronic kidney disease, malaria, and pneumonia as the specific diseases of interest. Chronic kidney disease is marked by a gradual decline in kidney function, leading to the accumulation of waste products in the blood and the onset of related health complications. On the other hand, malaria is a severe illness caused by Plasmodium parasites, which are transmitted through mosquito bites and can pose a significant threat to human life. Pneumonia, on the other hand, is an infection that inflames the air sacs in the lungs, often resulting in fever, fluid accumulation, and breathing difficulties. To facilitate rapid detection by lab assistants, our model leverages accurate disease prediction based on provided data. We fine-tune the neural network for chronic kidney disease prediction, while adopting convolutional neural networks (CNNs) for analyzing parasite-infected cells and chest X-rays in malaria and pneumonia detection, respectively. Additionally, we outline the potential expansion of this project to include other diseases in the future.

KEYWORDS: *Multi-disease detection, image classification, Deep Learning, Artificial Neural Network, Convolution neural networks.*

1. INTRODUCTION

Nowadays, most health system analysis models focus on analyzing individual diseases, lacking a unified system for disease detection. Consequently, health systems must train lab assistants to manually review patient reports and provide results. To address this issue, we propose a Flask API-based system capable of predicting multiple diseases. By analyzing and validating vast, intricate, and condensed data, this system facilitates knowledge extraction for informed decision-making. Lab assistants can effortlessly upload values and receive detection results, enhancing simplicity and accuracy across various hospitals.

This research will categorize the disease into infected and uninfected by identifying noteworthy patterns, recognising connections and linkages among several factors included in a vast dataset. It will be demonstrated as a significant source of precision and quick outcomes in the medical field. For even greater accuracy, the database will be automatically updated with any new data that is accurately anticipated. Using printing results as opposed to only providing results on the online site will help this project much further.

The diagnostic process typically involves the performance of one or more diagnostic procedures, which may include medical tests. It is crucial to accurately diagnose chronic illnesses due to the dependence on numerous variables that play a role in the diagnostic determination. It is a challenging technique

that frequently yields inaccurate conclusions. Humans are taught to identify diseases by only scanning values or visual cues. However, a computer must be programmed using a sizable dataset and by stating the relative relevance of the many accessible parameters. Experienced doctors can derive substantial benefits from the implementation of computer-aided diagnostic systems, particularly when aiming to render precise medical judgments. Consequently, there is significant interest within the medical community to automate the diagnostic process by fusing machine learning techniques with the expertise of physicians.

Machine learning technologies are increasingly employed to enhance diagnostic effectiveness by intelligently transforming available data. Numerous studies have explored the use of machine learning for diagnostics, demonstrating its ability to achieve 91.1% accuracy in comparison to highly skilled assistants, who achieve 79.97% accuracy.

2. LITERATURE SURVEY

The author of article [1] investigated several data mining algorithms for forecasting critical situations and renal illnesses. The author uses a dataset for chronic renal disease that has missing values. Missing values in the dataset reduce both the precision of the model and the forecast accuracy. They thus adjusted figures to fill up the gaps. The author developed a deep learning model [2] that uses a CNN [3] and a microscopic picture of infected blood cells to determine if an organism is infected with malaria. Using the presented model, which had a 95.23% accuracy rate, 15 out of 16 randomly generated pictures could be properly predicted. In a study described in reference [4], researchers explored the detection of pneumonia using a combination of machine learning models and mRMR (minimum Redundancy Maximum Relevance) feature selection technique on chest X-ray images. The outcomes of the feature extraction process using the existing CNN models, namely AlexNet, VGG16, and VGG19, are utilized as input and trained with five distinct ML techniques: Decision Trees, KNN (k-nearest neighbor), LDA (Linear Discriminant Analysis), LR (linear regression), and SVM (Support Vector Machines). With a 99.41% accuracy rate, LDA is the experiment's most accurate method. With the use of transfer learning, a model for detecting pneumonia illness in chest X-ray images was developed and published in [5]. This model builds a new architecture by combining pre-trained models [6] such as AlexNet, DenseNet121, InceptionV3, resNet18, and GoogleNet, and it achieves 96.39% accuracy on test pictures. According to certain parameters, a CNN-based classification model with transfer learning was created in [7] to detect pneumonia. The results were compared to identify the best model for the task. The author demonstrates how data normalisation affects the performance of the classification model in [8]. The main goal of this work was to investigate how data normalisation (scaling) affected the effectiveness of the classification model because the data set contained several numerical parameters with various scales. The illness, the parasite life cycle, the symptoms, the diagnosis, and the recommended course of therapy are all summarised in [9].

3. METHODOLOGY AND TECHNIQUES

In this study, we suggested a user-friendly, simple-to-understand system. By creating a more focused questionnaire that the system will use, we hope to reduce the amount of time needed. Our technology should serve as a channel for dialogue between users and lab assistants. The key to this will be machine learning and deep learning, and we'll employ techniques like convolutional neural networks (CNN), artificial neural networks (ANN), and logistic regression (LR) to produce accurate predictions.

The raw data from the original dataset is sent at the first stage, which is referred to as Data Pre-processing. During data pre-processing, duplicates, missing values, and other mistakes are eliminated from the raw data. A variety of computational models may be tested using the recently cleaned data.

The training of models' procedure is essential to any machine learning applications. Generally speaking, there are two machine learning methodologies: supervised learning and unsupervised learning. Our model initially predominantly employs the first technique, namely supervised learning. In supervised learning, the system is now taught on a set of examples (training set) before being asked to predict new values based on the test set.

Partitioning the dataset becomes necessary in order to get a model accuracy that is acceptable. The 80/20 rule is most typically applied when dividing. 80% of instruction and 20% of testing. We have employed two strategies—Logistic Regression and Artificial Neural Networks—for chronic renal disease. We have employed convolutional neural networks for malaria and pneumonitis.

3.1. Artificial Neural Network (ANN)

The idea behind an artificial neural network (ANN) is inspired by the remarkable capabilities of the human brain, which is a highly advanced system capable of analyzing real-world data. The fundamental building block of ANNs, as seen in Figure 1 [3], consists of a significant number of interconnected computing nodes (sometimes referred to as neurons), which work together to optimize the final output by learning from the input as a whole.

The input layer distributes the input to the hidden layers, which is normally loaded as a multidimensional vector. The learning process is when the hidden layers evaluate whether a stochastic change within themselves harms or improves the outcome in response to evaluations from the prior layer. Deep learning refers to systems that have several hidden layers constructed on top of one another.

$$y_i = bias + \sum_{i=1}^{i=n} w_i . x_i = w_1 . x_1 + w_2 . x_2 + \cdots + w_n . x_n \qquad (1)$$
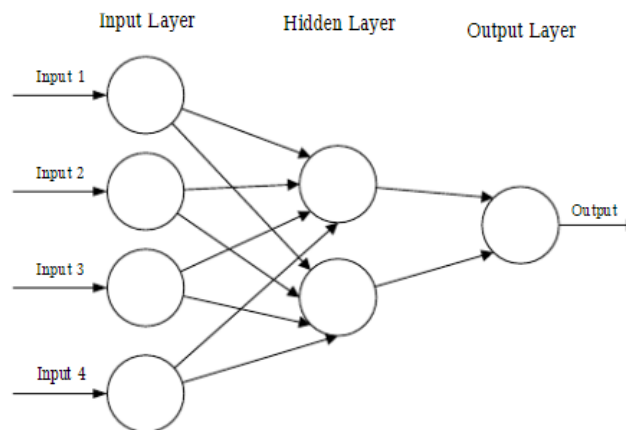

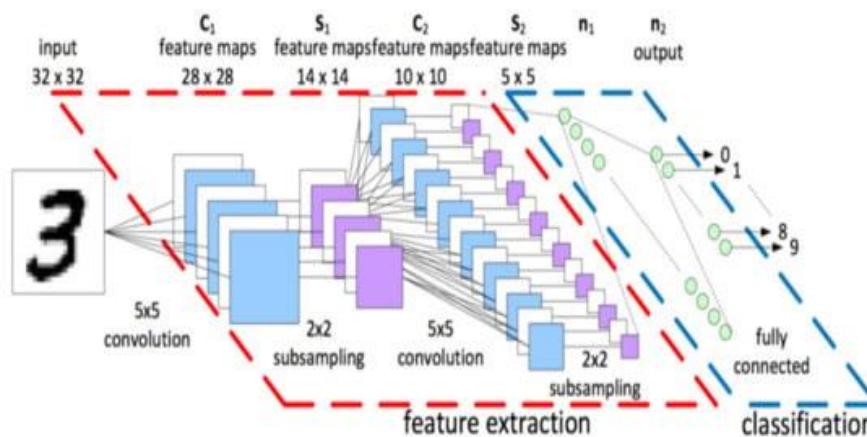**Figure. 1** A feedforward neural network with three layers

Figure 2: Convolutional neural network architecture

3.2. Convolution Neural Network (CNN)

The most popular tool for automatically interpreting visual images is CNN. One of the fundamental shortcomings of traditional ANN models is their inability to handle the computational complexity of image data. The architecture of a Convolutional Neural Network (CNN) comprises three main types of layers: input layers, convolutional layers, and dense layers. The fundamental components that enable the functioning of CNNs are weight sharing and the connections among neurons within each convolutional layer. Figure 2 [10] illustrates the overall architecture of a CNN model, which includes an input layer, convolutional layers with a ReLU activation function, pooling layers, and fully connected layers. This architecture is utilized for the identification of malaria and pneumonia.

3.3. Dataset Description

Out of the 25 features in the dataset for chronic kidney disease, we have chosen 13 crucial attributes needed to create a prediction model. Table 1 shows the value used for each attribute in chronic kidney dataset. Some attributes have Discrete Integer Values (DIV) and others have Nominal Values (NV): Normal, Abnormal. Table 2 presents the detail about pneumonia and malaria dataset.

Table 1: Chronic kidney dataset

Attributes	Value Used
Age	DIV
Blood Pressure	DIV
Albumin	DIV
Red Blood Cells	NV
Hypertension	NV
Serum creatinine	DIV
Sodium	DIV
Potassium	DIV
Haemoglobin	DIV
White blood cells count	DIV
Red blood cells count	DIV
Blood urea	DIV
Sugar	DIV

Table 2: Dataset for Pneumonia and Malaria

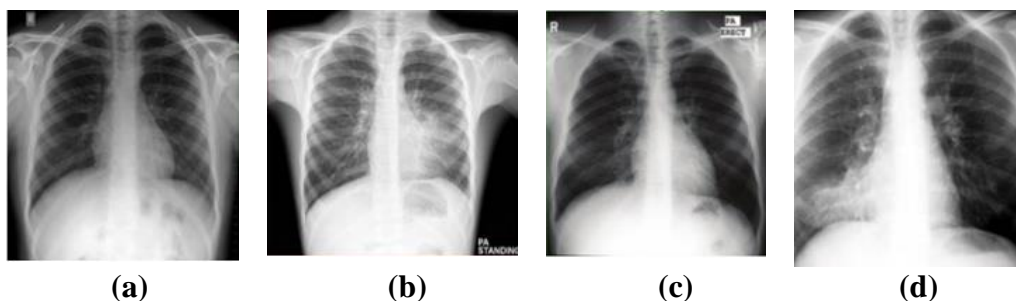| Disease	Normal	Infected
Malaria	Uninfected blood smear images	Parasitized blood smear images
Pneumonia (X-ray images)	Normal	Pneumonia

(a)	(b)	(c)	(d)

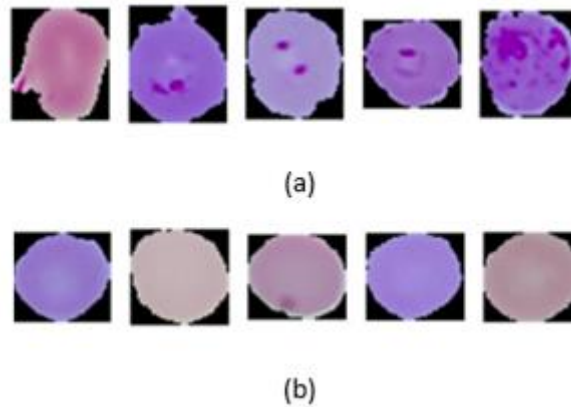**Figure: 3** (a and b)-Normal X-rays images, (c and d)- Pneumonia X-rays images.

Figure: 4. Blood smear images (a) Parasitized (b) Uninfected

4. CONCLUSION

In this paper, we present methods for recognising and categorising the many forms of sickness. We used the Kaggle and real-time dataset, and we employed attributes that are important to the illnesses. All the datasets are used for classification of disease. Artificial neural networks provide accuracy in the identification of chronic renal disease of about 98%. Convolutional neural networks are around 95% accurate in detecting malaria. This technique has the advantage of a speedier prediction process. Additionally, it will be simpler for doctors to diagnose more patients more quickly and to start treating patients sooner. This study contains constraints that prohibit it from being more robust due to the quantity of the data collection and the missing attribute values. The length of time required for the training step is determined by the complexity of the algorithm, which must be reduced as much as possible. More illnesses will be added to this system in the future. The authors hope that this research has aided the medical field in some manner and made the system more accessible to medical practitioners.

REFERENCES

[1] Arasu, S. D., & Thirumalaiselvi, R. (2017). Review of chronic kidney disease based on data mining techniques. International Journal of Applied Engineering Research, 12(23), 13498-13505.

[2] Gourisaria, M. K., Das, S., Sharma, R., Rautaray, S. S., & Pandey, M. (2020). A deep learning model for malaria disease detection and analysis using deep convolutional neural networks. International Journal of Emerging Technologies, 11(2), 699-704.

[3] O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

[4] M. Toğaçar, B. Ergen, Z. Cömert, F. Özyurt, "A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models," IRBM, S19590318193011742019, doi:10.1016/j.irbm.2019.10.006.

[5] V. Chouhan, S. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damasevicius, V. Albuquerque, "A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images," Applied Sciences, 10, 559, 2020, doi:10.3390/app10020559

[6] Marmanis, D., Datcu, M., Esch, T., and Stilla, U.: Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geosci. Remote Sens. Lett. 13(1), 105–109, (2016).

[7] Moujahid, H., Cherradi, B., Gannour, O. E., Bahatti, L., Terrada, O., & Hamida, S. (2020). Convolutional neural network based classification of patients with pneumonia using X-ray lung images. Advances in Science, Technology and Engineering Systems Journal, 5(5), 167-175.

[8] Borkin, D., Némethová, A., Michaľčonok, G., & Maiorov, K. (2019). Impact of data normalization on classification model accuracy. Research Papers Faculty of Materials Science and Technology IY Slovak University of Technology, 27(45), 79-84.

[9]   Tuteja, R. (2007). Malaria− an overview. The FEBS journal, 274(18), 4670-4679.

[10]   Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., ... & Ghayvat, H. (2021). CNN variants for computer vision: history, architecture, application, challenges and future scope. Electronics, 10(20), 2470.

AUTHORS

Manoj K. Singh received his BTech degree in information technology from UPTU (AKTU), Lucknow, India, in 2005. He received his MTech degree in computer science and engineering in 2010 from ISM (IIT) Dhanbad, India. He received his PhD degree in computer science and information technology in May 2023 from MJP Rohilkhand University, Bareilly, India. Currently he is working as an Assistant professor in CSE department of MIT Moradabad, UP, India. He has more than 17 years of teaching experience and has published more than 20 research papers. His research interests include remote sensing, image processing, hyperspectral image classification, machine learning and deep learning.

Simran Arora completed her B.Tech in computer science and Engineering in 2023 from MIT Moradabad, UP, India. Her research interests include machine learning, deep learning and image processing.

Shobhit Saini completed his B.Tech in computer science and Engineering in 2023 from MIT Moradabad, UP, India. His research interests include machine learning, deep learning and image processing.

Wazahat Ali completed his B.Tech in computer science and Engineering in 2023 from MIT Moradabad, UP, India. His research interests include machine learning, deep learning and image processing.

Faraz Jafri completed his B.Tech in computer science and Engineering in 2023 from MIT Moradabad, UP, India. His research interests include machine learning, deep learning and image processing.