A Fusion of Multiple Neural Networks Based on Machine Learning and Deep Learning for Multiple Disease Detection

Manoj Kumar Singh^{#1}, Simran Arora^{#2}, Shobhit Saini^{#3}, Wazahat Ali^{#4}, Faraz Jafri^{#5} [#]Department of Computer Science & Engineering Moradabad Institute of Technology, Moradabad, UP, India ¹manojaswal1982@gmail.com, ²simran.arora1271@gmail.com, ³shobhitsaini029@gmail.com, ⁴Official8ali@gmail.com, ⁵Jafrifaraz4@gmail.com.

Abstract- In today's time, people want everything fast. When analyzing medical data, imbalanced data is a significant issue. This paper focuses on the problem of lab assistants to precisely look into the values and report of patients to predict the occurrence of disease. In our paper, we proposed a supervised model for detecting multiple diseases. Our model is based on machine learning and deep learning techniques. Deep learning and machine learning Chronic-kidney diseases, Malaria diseases, Pneumonia diseases. Chronic kidney disease does a gradual loss of kidney functions. It decreases their ability to keep an individual healthy by filtering wastes from your blood. If kidney disease worsens, wastes can build to high levels in blood and make feel sick. Malaria can be a life-threatening disease caused by Plasmodium parasites that are spread to people through the bites of infected female Anopheles mosquitoes. Pneumonia is a type of infection that causes the air sacs in one or both lungs to expand and become inflamed. This may result in fever, pus or fluid in the lungs of any person, and breathing difficulties. To help lab assistants for quick detection, we are focusing on providing the accurate disease prediction with the data provided. For Chronic-kidney, we tuned finely the neural network for better performance in forecasting the likelihood of kidney disease. For Malaria and Pneumonia, CNN is used on the images of parasite-infected cells and chest X-rays respectively. Adding further disease is the future scope of this project.

Keywords— Classification, hyperspectral image, remote sensing, convolution neural networks.

I. INTRODUCTION

These days many of the existing learning models for health system analysis are concentrating on an individual disease per analysis. There is no common system that can detect the disease, health systems have to train lab assistants to check the reports of the patients and give the results. In this we are proposing a system which is used to predict multiple diseases by using Flask API. To extract knowledge for better decisions, the huge, complex, and compact data must be analyzed and verified. Lab assistants will simply upload the values and after submitting will get the detection result. This will provide simplicity and accuracy to the lab assistants of different hospitals.

This paper identifies significant patterns and detecting correlations and relationships among many variables present in enormous dataset and will classify the disease into infected and uninfected. It will be proven as an important source of accuracy and fast yield of results in medical sector. Any new data that is predicted correctly will be automatically added to the database for further better accuracy. This project can be further improved by using printable result instead of showing results only on the web application.

One or more diagnostic procedures, such as medical tests, are carried out during the diagnostic process. The diagnosis of chronic diseases is necessary since it depends on numerous factors. It is a difficult process that frequently results in false assumptions. Human are trained to detect the disease by simply looking at the value or image. But computer has to be trained with huge dataset and by specifying the importance ratio of different parameters available. A computer-aided diagnostic system can aid even seasoned doctors in rendering reliable medical decisions. Therefore, combining machine learning techniques with physician experience to automate the diagnosing process is of great interest to the medical community.

In order to increase the effectiveness of the diagnostic process, machine learning technologies are making major efforts to intelligently convert accessible data into useful information. To investigate the utilisation of machine learning in terms of diagnostic capabilities, several studies have been carried out. It was shown that machine learning algorithms could identify with 91.1% accuracy when compared to the most skilled assistants, who can diagnose with 79.97% accuracy.

II. LITERATURE SURVEY

In paper [1] author looked into different data mining techniques for predicting serious issues and kidney diseases. The author makes use of a dataset with missing values for chronic kidney disease. The precision of the model and the accuracy of the predictions are both lowered by missing values in the dataset. As a result, they recalculated values to replace the missing ones. In order to detect whether an organism has malaria, the author created a deep learning model [2] that makes use of a CNN [3] and a microscopic image of blood cells that are infected with the disease. 15 out of 16 randomly generated photos could be correctly predicted using the provided model, which had an accuracy rate of 95.23%. In [4], the authors published a study on the diagnosis of pneumonia using a mix of machine learning models and mRMR feature selection on X-ray images of the chest. Decision trees, k-nearest neighbors, linear discriminant analysis, linear regression, and support vector are five alternative machine learning models that are fed and trained

using the outcomes of the features extraction process using the current CNN models AlexNet, VGG16, and VGG19. The experiment yields the highest accuracy of 99.41% for linear discriminant analysis. A model created for chest X-ray images of pneumonia disease using the transfer learning approach was released in [5]. This model generates a novel architecture by merging pre-trained models AlexNet, DenseNet121, InceptionV3, resNet18, and GoogleNet, and it achieves 96.39% accuracy on test images. A CNN-based classification model with transfer learning was developed in [6] to diagnose pneumonia and the findings were compared to determine the optimal model for the job depending on particular criteria. In [7], the author shows how data normalisation influences the classification model's efficacy. Because the data set contains numerous numerical parameters with different scaling, the primary objective of this work was to explore the impact of data normalisation (scaling) on the efficacy of the classification model. In [8] the author summarises the sickness, the parasite life cycle, the symptoms, the diagnosis, and the course of treatment.

III. METHODOLOGY AND TECHNIQUES

In this work, we proposed a user-friendly system that is easy to learn and utilize. We are attempting to shorten the time needed by developing a more targeted questionnaire that the system will employ. We want our system to act as a communication link between users and lab assistants. In this, Machine learning and Deep learning will be the key component, and to achieve this, we'll use methods like convolutional neural networks (CNN), artificial neural networks (ANN), and logistic regression (LR), which will allow us to make predictions that are precise.

The first phase, known as Data Pre-processing, is where the raw data from the original dataset is transferred. Redundancies, missing values, and other errors are removed from the raw data during data pre-processing. The newly cleaned data can be used to test various computational models.

Fundamental to any machine learning projects is the training of model process. In general, *supervised learning* and *unsupervised learning* are the two machine learning methodologies. The first strategy is primarily used by our model at beginning, that is, Supervised Learning. Now, in supervised learning, the model is asked to predict new values based on the test set after the system has been trained on a collection of instances (training set).

In order to get acceptable model accuracy, dataset partitioning becomes essential. When dividing, the 80/20 rule is most frequently utilised. 20% for testing and 80% for training. For Chronic kidney disease, we have used two approaches – Logistic Regression and Artificial Neural Network. For Malaria and Pneumonia, we have used Convolutional Neural Network.

A. Artificial Neural Network (ANN)

The idea of an artificial neural network (ANN) is based on the notion that the human brain is a more intelligent system capable of processing real data. The basic component of ANNs as shown in the Figure 1 [3] is a large number of connected computational nodes (also known as neurons), which collaborate to optimize the final output by learning from the input as a whole.

The input, which is typically loaded as a multidimensional vector, is distributed to the hidden layers by the input layer. Following judgements from the preceding layer, the hidden layers will evaluate whether a stochastic change inside itself worsens or improves the result, and this is known as the learning process. Deep learning is a term used to describe systems with many hidden layers built on top of one another.

$$y_i = bias + \sum_{i=1}^{l=n} w_i \cdot x_i = w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n$$
(1)





Figure. 1 Three layered feedforward neural network

Figure 2: Convolutional neural network architecture

B. CONVOLOTION NEURAL NETWORK (CNN)

CNN are most frequently used to interpret visual pictures automatically. Traditional ANN models struggle with the computational complexity needed to calculate picture data, which is one of their main drawbacks. Input layers, convolutional layers, and dense layers are the three different types of layers that make up CNN's architecture. Sharing weights and connections between the neurons in each convolutional layer are typically the two fundamental components on which convolutional neural networks base their functionality. Figure 2 [9] shows the general architecture of a CNN model used to detect malaria and pneumonia, containing an input layer, convolution layers with ReLU activation function, pooling layers, and fully connected layers.

C. DATASET DESCRIPTION

For Chronic kidney disease, out of 25 attributes present in the dataset, we have selected 13 important attributes required to build predictive model.

Attributes	Value Used	
Age	Discrete Integer Values	
Blood Pressure	Discrete Integer Values	
Albumin	Nominal Values	
Red Blood Cells	Nominal Values (Normal, Abnormal)	
Hypertension	Nominal Values (Normal, Abnormal)	
Serum creatinine	Discrete Integer Values	
Sodium	Discrete Integer Values	
Potassium	Discrete Integer Values	
Haemoglobin	Discrete Integer Values	
White blood cells count	Discrete Integer Values	
Red blood cells count	Discrete Integer Values	
Blood urea	Discrete Integer Values	
Sugar	Discrete Integer Values	

Table 1: Chronic kidney dataset

Figure 3 [6] shows few dataset images for Malaria and Pneumonia. Fig. 3a and Fig. 3b show the normal lung X-rays, while Fig. 3c and Fig. 3d show the pneumonic lung X-rays. Similarly, in Figure 4 [2], upper blood smear cells show parasitized cells and other is showing uninfected one.

Table 2: Dataset used for Malaria and Pneumonia

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



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.

IV. CONCLUSION

In this paper we have demonstrated approaches to identify and classify the various types of disease. We have used the Kaggle and real-time dataset and specific features are used relevant to the diseases. All the datasets are used for classification of disease. In Chronic kidney disease detection, Artificial neural network gives approximately 98% accuracy. In Malaria disease detection, Convolutional neural network gives approximately 95% accuracy. The prediction process is quicker with this method, which is a benefit. Additionally, it will make it easier for the medical professionals to diagnose more patients in a shorter amount of time and to begin treating patients earlier. Due to the size of the data collection and the missing attribute values, this study has certain limitations that prevent it from being more robust. The length of time required for the training stage depends on the complexity of the algorithm that has to be as well optimised as feasible. The future of this system will be the addition of more diseases. The authors hope that this paper has in some way helped the medical field and made the system more accessible to the medical professionals.

REFERENCES

- 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] 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.
- [7] Borkin, D., Némethová, A., Michal'čonok, G., & Maiorov, K. (2019). Impact of data normalization on classification model accuracy. Research Papers Faculty of Materials Science and Technology Slovak University of Technology, 27(45), 79-84.
- [8] Tuteja, R. (2007). Malaria- an overview. The FEBS journal, 274(18), 4670-4679.
- [9] 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.