International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| 9th ICAILLMBA (ALLAM) 2025-Conference Proceedings
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.ICAILLMBA-105

Original Article

Automated Prediction of Multiple Diseases through Deep Learning Models

Malla Sireesha[1] ,Ch. Manikanta Kalyan[2], Gantyada AmalaDevi[3], Priyanth Dunna[4]
[1]Department of Information Technology and Computer Applications, Andhra University, Visakhapatnam, Andhra Pradesh, India
[2]Department of AI & ML, Aditya University, Andhra Pradesh, India
[3]Assistant Professor, Department of CSE, Visakha Institute of Engineering & Technology, Visakhapatnam, India
[4]Department of Artificial Intelligence and Machine Learning, Aditya University, Surampalem, Kakinada, Andhra Pradesh, India

Abstract *- Accurate and early disease prediction is essential in modern healthcare due to the rising prevalence of chronic illnesses. This project presents a deep learning framework for predicting multiple diseases designed to improve diagnostic accuracy and efficiency using advanced neural network models. The system applies effective preprocessing techniques, including normalization and Z-score scaling, to enhance feature representation of critical health parameters such as glucose, hemoglobin, cholesterol, and blood pressure. A Multi-Layer Perceptron (MLP) model is employed because of its capacity to learn intricate non-linear patterns while maintaining a simpler and efficient architecture compared to deep feedforward networks. Model training is performed with the Adam optimizer and validated through cross-validation to ensure robustness across diverse patient data, resulting in an accuracy of 95.13%. The proposed system is implemented using Python with TensorFlow, pandas, NumPy, and scikit-learn libraries on a Windows platform.*

Keywords *- Multi-Disease Prediction, Deep Learning, Multi-Layer Perceptron, Blood Parameter Analysis, Healthcare Analytics.*

1. Introduction

In recent years, the rapid increase in chronic and lifestyle-related diseases has posed significant challenges to healthcare systems worldwide. Early and accurate disease prediction plays a vital role in enhancing patient outcomes, lowering healthcare costs, and supporting timely medical intervention. Blood test parameters such as glucose levels, hemoglobin, cholesterol, and blood pressure provide valuable clinical insights and are widely used for diagnosing multiple diseases. However, traditional diagnostic approaches often rely on manual analysis and disease-specific models, which may limit their effectiveness when dealing with complex, multi-parameter medical data.

Within the field of artificial intelligence, deep learning methods are increasingly used in medical diagnosis because of their capacity to extract complex patterns from large-scale data. Existing systems employ models such as CNNs, RNNs, and ensemble learning techniques for disease prediction. Although these approaches have demonstrated promising results for specific conditions such as diabetes and heart disease, their performance varies across different medical conditions. Moreover, deep learning and ensemble models often involve high computational complexity, making them difficult to interpret, maintain, and deploy in real-time clinical environments.

The variability in prediction accuracy across diseases and the complexity of existing models highlight the need for a more efficient and unified solution for multi-disease prediction. An ideal system should be capable of accurately predicting multiple diseases from blood sample data while maintaining a simple and scalable architecture. Additionally, such a system should ensure consistent performance, reduced computational overhead, and improved generalization across diverse patient datasets.

This research proposes a deep learning–driven multi-disease prediction framework developed to handle these limitations by applying appropriate data preprocessing strategies, robust feature representation, and a Multi-Layer Perceptron (MLP) model, the proposed approach aims to predict multiple diseases with a high level of accuracy and robustness. The system focuses on improving diagnostic efficiency and supporting medical practitioners by enabling early disease identification and effective treatment planning, leading to improved patient management and healthcare decisions.

2. Related Work

Qiu et al. proposed a clinical knowledge-aware framework for multi-disease prediction that integrates medical knowledge with machine learning models, achieving improved prediction accuracy compared to purely data-driven approaches. However, the method depends on the availability of structured clinical knowledge, which may limit its scalability. The authors conclude that incorporating clinical expertise enhances multi-disease predictive performance [1]. Yaganteeswarudu et al. developed a machine learning-based framework for multi-disease prediction using a Flask API to provide a unified and accessible

healthcare prediction system. The model demonstrated reasonable accuracy in predicting multiple diseases using symptom-based inputs. However, its performance is limited by dependency on predefined datasets and lack of real-world clinical validation. The study concludes that integrating machine learning with web frameworks enhances scalable and user-friendly disease prediction systems [2].

Park et al. developed an ensemble machine learning model using laboratory test data to predict multiple diseases with high accuracy. The model showed strong predictive performance and improved interpretability through feature analysis. However, it relies on retrospective clinical data, which may limit real-world generalization. The study concludes that machine learning models based on routine lab tests can effectively support clinical decision-making [3]. Men et al. developed a deep learning framework utilizing LSTM networks model for multi-disease prediction using longitudinal clinical records, effectively capturing temporal dependencies among patient visits. The model achieved superior prediction performance compared to traditional methods. However, it requires large-scale sequential data and has high computational complexity, limiting practical deployment. The study concludes that temporal deep learning models are effective for multi-disease clinical decision support [4].

Rehman et al introduced a machine learning–driven multi-disease prediction model that analyzes patient medical data to predict diseases including diabetes and cardiac disorders, liver, and kidney disorders. The model showed effective prediction performance, demonstrating the potential of ML for early disease detection. However, limitations include data dependency and challenges in real-world clinical deployment. The study concludes that integrated machine learning systems can support proactive and efficient healthcare diagnosis [5]. Bharath et al. developed a machine learning-enabled multi-disease prediction framework utilizing SVM, Random Forest and Decision Tree classifiers to analyze clinical data with high accuracy, with SVM showing superior performance. However, the approach depends on selected features and limited datasets, affecting generalization. The study concludes that machine learning can effectively support early multi-disease diagnosis [6].

Mali et al. presented a machine learning-based multiple disease prediction system using classification algorithms to analyze health data for early diagnosis. The model showed improved prediction performance through feature selection and ensemble techniques. However, its accuracy depends on data quality and dataset diversity. The study concludes that machine learning can effectively support multi-disease prediction in healthcare [7]. Ampavathi et al. introduced a hybrid deep learning–based multi-disease prediction framework using optimized feature extraction and neural networks. The model achieved improved prediction accuracy across multiple diseases. However, its complex architecture increases computational cost and limits real-world scalability. The findings indicate that well-tuned deep learning models can significantly improve multi-disease healthcare prediction [8].

Chunduru et al. developed a hybrid model that uses CNN-based feature representation followed by Random Forest classification to predict chronic diseases. The hybrid model achieved strong prediction performance across multiple diseases. However, its reliance on specific datasets and higher computational cost may limit generalizability. The study concludes that CNN–Random Forest integration improves multi-disease prediction accuracy [9]. Zheng et al. developed a framework based on multi-modal graph learning to automatically learns latent graph structures and integrates multiple data modalities for disease prediction, achieving superior performance on benchmark prediction tasks. The model effectively captures modality correlations and enhances patient representation learning. However, its complexity and dependency on multi-modal data may limit applicability in settings with incomplete data. The study concludes that adaptive multi-modal graph learning improves accuracy and robustness in disease prediction tasks [10].

3. Methodology

3.1. Data Collection

A labeled medical dataset is collected comprising patient clinical records with features such as glucose concentration, hemoglobin levels, cholesterol values, blood pressure, red blood cell count, white blood cell count, and platelet count, along with corresponding disease representing diabetes, anemia, thalassemia, cardiovascular disorders, thrombocytopenia, and individuals without diagnosed conditions

The proposed multi-disease prediction framework is implemented using a Multi-Layer Perceptron (MLP) architecture, as illustrated in the given figure 1. The methodology follows a systematic deep learning pipeline to accurately classify multiple diseases using blood sample data.

The dataset utilized in this research comprises blood parameter measurements associated with multiple disease conditions. Prior to model development, the data undergoes a preprocessing phase to ensure quality and consistency. Missing values are appropriately handled, and all numerical features are normalized using Z-score normalization to standardize feature ranges. This preprocessing step enhances training stability, accelerates convergence, and improves overall learning efficiency. The preprocessed blood parameters are then provided to the input layer of the neural network in the form of feature vectors. Each

neuron in the input layer represents a specific clinical attribute, ensuring that all relevant medical information is effectively conveyed to the model for learning disease-related patterns.

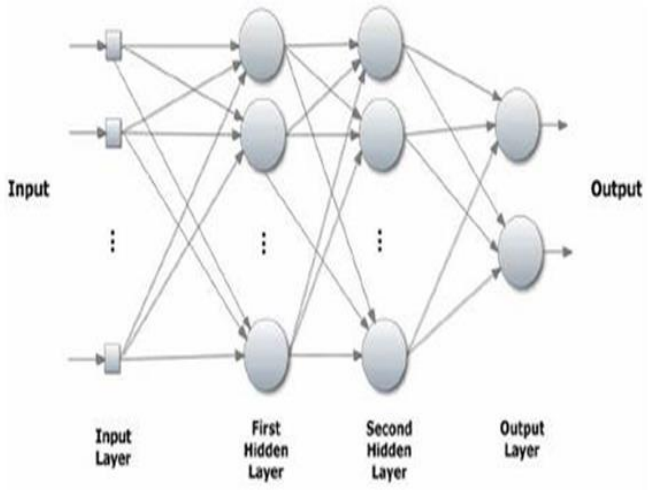

Fig 1: MLP Architecture

The network architecture includes multiple fully connected hidden layers, as depicted in the model design. Within these layers, neurons compute weighted sums of inputs, incorporate bias terms, and apply ReLU, a commonly used non-linear activation function. This hierarchical structure enables the model to capture complex and non-linear relationships among blood parameters, thereby improving its capability to differentiate among various the network further enhances feature abstraction and contributes to higher classification accuracy. The output layer consists of neurons equal to the number of disease classes. A softmax activation function is applied to generate probability distributions for all disease classes, with the label corresponding to the highest probability chosen as the final diagnosis.

To prepare the model for training, it is compiled using the Adam optimization algorithm, which efficiently updates network weights during backpropagation. For multi-class prediction, the loss is computed using categorical cross-entropy while accuracy is used as the primary performance metric. Model training is conducted over multiple epochs using the training dataset. During this phase, training and validation accuracy and loss are consistently tracked to assess learning behavior and mitigate overfitting. To reduce the impact of class imbalance, oversampling approaches are adopted, ensuring balanced learning across all disease classes.

After the training phase, model performance is measured using an independent test dataset to assess its generalization capability. Performance is measured using accuracy, precision, recall, F1-score, and detailed classification reports. Additionally, graphical analyses of accuracy and loss across training, validation, and test phases provide further insight into the model's robustness and stability. For prediction, new blood sample data is passed through the trained network to generate disease classifications. Experimental results demonstrate that the MLP-based model outperforms other approaches, making it an effective and reliable solution for early and accurate multi-disease prediction.

4. Results

An extensive performance analysis of the developed multi-condition prediction framework developed using blood parameter data. The performance of the trained models is evaluated using standard classification measures. A comparative analysis is performed across multiple architectures, including MLP, DFNN, CNN, and RNN, to determine the most suitable model for disease prediction. Furthermore, performance evaluation using accuracy and loss for training, validation, and test sets curves across epochs are analyzed to study the learning characteristics, generalization ability, and overfitting behavior of each model. The experimental findings offer meaningful insights into the reliability and effectiveness of the proposed approach for accurate multi-disease classification.

Classification Report:

	precision	recall	f1-score	support
Healthy	0.06	0.40	0.11	5
Disease	0.99	0.94	0.96	481
accuracy			0.93	486
macro avg	0.53	0.67	0.54	486
weighted avg	0.98	0.93	0.96	486

Fig 2: Classification Report of the Trained Model

The above Figure 2 presents the classification report generated after training the model, presenting essential performance metrics for each disease category. These evaluation measures provide an in-depth analysis of model effectiveness, highlighting how well the model predicts each disease and identifying areas for potential improvement.

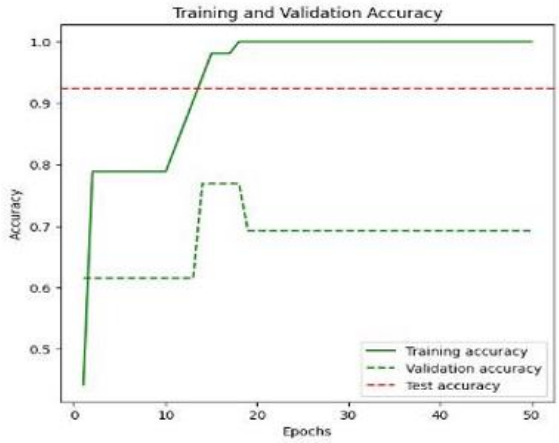

Fig 3: Plot diagram of Training,Validation and Test Accuracy

The above Figure 3 plot shows the model's accuracy over epochs with training accuracy (solid green line), validation accuracy (dashed green line), and test accuracy (red dashed horizontal line). It visualizes learning trends, model performance on unseen data, and generalization capability. The comparison helps assess the effectiveness and overfitting of the model.

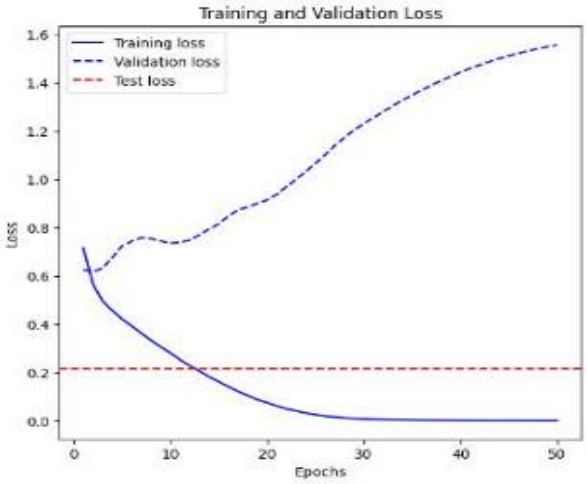

Fig 4: Plot diagram of Training,Validation and Test Loss

The above Figure 4 plot displays the model's loss over epochs with training loss (solid blue line), validation loss (dashed blue line), and test loss (red dashed horizontal line). It illustrates how the model's loss decreases during training, monitors performance on validation data, and evaluates generalization on test data. The comparison highlights the model's learning progress and potential overfitting.

Table 1: Model Performance Comparison

SL.No	Model	Accuracy
1	**MLP Model**	**95.13%**
2	DFNN Model	83.84%
3	CNN Model	78.50%
4	RNN Model	85.67%

The above table 1 presents the accuracy achieved by different machine learning approaches to identify diseases from blood parameters and the proposed model (i.e.,MLP Model) got high accuracy compared with alternative models. This visualization provides a clear comparison of the models' effectiveness, highlighting which model performed best in terms of accuracy.

5. Conclusion

This study effectively illustrates the capability of modern deep learning approaches, especially Feed forward Neural Networks and Multi-Layer Perceptrons optimized using the Adam optimizer, for multi-disease prediction from blood sample data. The developed models accurately classified diseases such as Diabetes, Anemia, Thalassemia, Heart Disease, and Thrombocytopenia, highlighting their potential applicability in medical diagnostics. Challenges such as class imbalance were addressed using oversampling techniques, leading to improved predictive performance. The experimental results indicate that deep learning models hold significant promise for early disease detection and clinical decision support. Looking ahead, further advancements can be achieved by incorporating larger and more diverse datasets, improving model interpretability, ensuring robustness across populations, and integrating genomic, lifestyle, and environmental factors. With careful attention to data privacy and bias, such systems can contribute to more accurate diagnoses, personalized treatment strategies, improved healthcare efficiency, and better patient outcomes.

References

[1] L. Qiu, S. Gorantla, V. Rajan, and B. C. Y. Tan, "Multi-disease predictive analytics: A clinical knowledge-aware approach," ACM Trans. Manage. Inf. Syst., vol. 12, no. 3, Art. no. 19, pp. 1–34, Sep. 2021, doi: 10.1145/3447942.

[2] 2. Akkem, Yaganteeswarudu. (2020). Multi Disease Prediction Model by using Machine Learning and Flask API. 1242-1246. 10.1109/ICCES48766.2020.9137896

[3] Park, D.J., Park, M.W., Lee, H. et al. Development of machine learning model for diagnostic disease prediction based on laboratory tests. Sci Rep **11**, 7567 (2021). https://doi.org/10.1038/s41598-021-87171-5

[4] Lu, Men & Ilk, Noyan & Tang, Xinlin & Liu, Yuan. (2021). Multi-Disease Prediction Using LSTM Recurrent Neural Networks. Expert Systems with Applications. 177. 114905. 10.1016/j.eswa.2021.114905.

[5] Rehman, Ahad. (2024). Multiple Disease Prediction System using Machine Learning. Journal of Informatics Electrical and Electronics Engineering (JIEEE). 5. 1-13. 10.54060/a2zjournals.jieee.110.

[6] C. Bharath *et al.*, "Multi Disease Prediction using Machine Learning Algorithms," Int. J. Innov. Res. Sci., Eng. Technol., vol. 13, no. 6, Jun. 2024.

[7] Mali, Kshitija and Rasane, Pooja and Patil, Mohini and Suryawanshi, Gayatri and Chaphadkar, Anagha, Multiple Disease Prediction using Machine Learning (Feb 27, 2024). Available at SSRN: https://ssrn.com/abstract=4747365 or http://dx.doi.org/10.2139/ssrn.4747365

[8] Ampavathi A, Saradhi TV. Multi disease-prediction framework using hybrid deep learning: an optimal prediction model. Comput Methods Biomech Biomed Engin. 2021 Aug;24(10):1146-1168. doi: 10.1080/10255842.2020.1869726. Epub 2021 Jan 11. PMID: 33427480.

[9] Chunduru, A., Kishore, A.R., Sasapu, B.K. *et al.* Multi Chronic Disease Prediction System Using CNN and Random Forest. SN COMPUT. SCI. **5**, 157 (2024). https://doi.org/10.1007/s42979-023-02521-6

[10] S. Zheng et al., "Multi-Modal Graph Learning for Disease Prediction," in IEEE Transactions on Medical Imaging, vol. 41, no. 9, pp. 2207-2216, Sept. 2022, doi: 10.1109/TMI.2022.3159264.