IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

A Review on Deep Learning-Based AI Systems for Clinical Decision Support and Disease Prediction

Vishal Vishnu Gurav[1] and Prof. D. B. Thakur[2]
Department of Electronics & Telecommunication[1]
Associative Professor, Department of Electronics & Telecommunication[2]
TPCT's College of Engineering, Dharashiv, India

Abstract: *The exponential growth of healthcare data has created vast opportunities for intelligent disease prediction, which plays a crucial role in improving patient care and reducing treatment costs. This research presents a smart disease prediction and health assistance system powered by the XGBoost machine learning model, integrating patient medical history and symptom analysis. The proposed system accepts user-input symptoms, processes them through the trained XGBoost classifier, and predicts the most probable disease with high accuracy. Furthermore, it categorizes the disease severity into low, moderate, high, or extreme levels. Based on the prediction and severity analysis, the system provides personalized health recommendations—including suitable medications, diet plans, exercise routines, and necessary precautions—while advising medical consultation for severe conditions. By leveraging structured healthcare datasets and advanced machine learning techniques, this approach enhances prediction precision, promotes preventive healthcare, and improves accessibility for patients in remote regions. Ultimately, the system bridges the gap between automated diagnosis and personalized medical guidance, contributing to a more efficient, data-driven healthcare ecosystem.*

Keywords: XGBoost, Disease Prediction, Machine Learning, Symptom Analysis, Medical Recommendation System, Patient Management

I. INTRODUCTION

Healthcare is one of the most critical domains where technological advancements have a direct and profound impact on human life. With the exponential growth of healthcare data from electronic health records (EHRs), wearable devices, and diagnostic tools, the need for accurate analysis and disease prediction has become increasingly significant. Traditional medical diagnosis typically relies on expert consultation, extensive laboratory testing, and considerable processing time. These limitations often restrict timely access to quality healthcare, particularly for patients residing in remote or underserved regions.

In this context, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative technologies capable of reshaping the healthcare landscape. By analyzing vast amounts of medical data, these intelligent systems can assist in disease prediction, risk assessment, and personalized health recommendations. Among various ML techniques, models such as Extreme Gradient Boosting (XGBoost) have demonstrated exceptional performance in classification and predictive analytics across multiple domains, including healthcare. Leveraging such models enables the development of intelligent systems capable of accurately predicting diseases based on patient symptoms, medical history, and physiological parameters.

The proposed system utilizes patient-reported symptoms as input and applies XGBoost-based predictive modeling to identify the most probable disease. Additionally, it assesses the severity level of the predicted disease by categorizing it into low, moderate, high, or extreme classes according to weighted symptom importance. Based on these outcomes, the system generates personalized recommendations, including suggested medications, dietary plans, exercise routines, and preventive precautions. In cases where severe or critical conditions are detected, the system alerts the user to seek immediate medical attention from a qualified healthcare professional.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, November 2025 Impact Factor: 7.67

By integrating structured healthcare data analysis with machine learning techniques, the proposed approach aims to support both patients and healthcare providers. For patients, it offers an accessible, low-cost, and instant health assessment platform, while for medical practitioners, it functions as an intelligent clinical decision-support tool that enhances diagnostic accuracy and efficiency. Furthermore, the system incorporates patient management modules, digital health records, and dedicated doctor panels to facilitate continuous monitoring and care. Overall, this research contributes to the development of a smart, preventive, and personalized healthcare framework that leverages AI to promote early diagnosis, improve treatment outcomes, and advance the vision of next-generation intelligent healthcare systems.

II. LITERATURE SURVEY

The prediction of disease at earlier stage becomes important task. But the accurate prediction on the basis of symptoms becomes too difficult for doctor. There is a need to study and make a system which will make it easy for end users to predict the chronic diseases without visiting physician or doctor for diagnosis. Table 1 shows literature survey about disease prediction systems proposed in different literatures.

Table 1 literature review

Sr. no.	Paper Name, Author and year	Outline	Advantages
1	A Medical-History-Based Potential Disease Prediction Algorithm, Wenxing et al, IEEE Access	This paper proposed novel deep-learning-based hybrid recommendation algorithm, which predicts the patient's possible disease based on the patient's medical history and provides a reference to patients and doctors	1) It considers both, high-order relations as well as low order combination of disease among disease features, 2) Improved comprehensiveness compared to previous system.
2	Designing Disease Prediction Model Using Machine Learning Approach, Dahiwade, D., Patle, G., & Meshram, E., IEEE Xplore/	Proposed general disease prediction, In which the living habits of person and checkup information consider for the accurate prediction It also computes the risk associated with general disease	1) low time consumption 2) minimal cost possible 3) The accuracy of disease prediction is 84.5%
3	Explainable Learning for Disease Risk Prediction Based on Comorbidity Networks, Xu, Z., Zhang, J., Zhang, Q., & Yip, P. S. F., IEEE/	Proposed a comorbidity network involved end-to-end trained disease risk prediction model. The prediction performances are demonstrated by using a real case study based on three years of medical histories from the Hong Kong Hospital Authority.	1) Comfortably incorporates the comorbidity network into a Bayesian framework 2) Exhibits superior prediction performance
4	Design And Implementing Heart Disease Prediction Using Naives Bayesian, Repaka, A. N., Ravikanti, S. D., & Franklin, R. G., IEEE/	This paper focused on heart disease diagnosis by considering previous data and information. To achieve this SHDP (Smart Heart Disease Prediction) was built via Navies Bayesian in order to predict risk factors concerning heart disease.	1) Accuracy is 89.77% in spite of reducing the attributes. 2) The performance of AES is highly secured compared to previous encrypting algorithm (PHEC).
5	Similar Disease Prediction with Heterogeneous Disease	Proposed a method to predict the similarity of diseases by node representation learning.	1) As the range of predictions expands, the proposed method is better than the disease prediction of

	Information Networks, Gao, J., Tian, L., Wang, J., Chen, Y., Song, B., & Hu, X., IEEE/		only chemical-disease data source
6	Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning, Mathew, R. B., Varghese, S., Joy, S. E., & Alex, S. S., IEEE/	This paper explained a medical chatbot which can be used to replace the conventional method of disease diagnosis and treatment recommendation. Chatbot can act as a doctor.	1) This system help in reducing conduction of daily check-ups 2) It identifies the symptoms and gives proper diagnosis. 3) Chatbot doesn't require the help of physician 4) Cheaper 5) The chat and users relation is completely personal which helps users to be more open with their health matters
7	Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning, Maurya, A., Wable, R., Shinde, R., John, S., Jadhav, R., & Dakshayani, R., IEEE/	The proposed system use machine learning algorithm and suggest suitable diet plan for CKD patient using classification algorithm on medical test records. This extracts the features which are responsible for CKD, then machine learning process can automate the classification of the chronic kidney disease in different stages according to its severity.	1) Detects and suggest diet which will be useful to the doctors as well as patients
8	Designing Disease Prediction Model Using Machine Learning Approach, Dahiwade, D., Patle, G., & Meshram, E., IEEE/	This system compares CNN and KNN for disease prediction Disease dataset from UCI machine learning website is extracted in the form of disease list and its symptoms. Pre-processing is performed on that dataset. After that feature extracted and selected. Then classification and prediction using KNN and CNN is performed.	1) The CNN takes less time than KNN for classifying large dataset. 2) CNN gives more accurate disease prediction than KNN.
9	Smart Health Monitoring System using IOT and Machine Learning Techniques, Pandey, H., & Prabha, S., IEEE/	This paper deal with IoT which helps to record the real time (patient) data using pulse rate sensor and arduino and is recorded using thing speak. Machine learning algorithms were used to make prediction of heart disease.	1) The proposed system helps patient to predict heart disease in early stages. 2) It will be helpful for mass screening system in villages where hospital facilities are not available.
10	Random Forest Algorithm for the Prediction of Diabetes, VijiyaKumar, K., Lavanya, B., Nirmala, I., & Caroline, S. S, IEEE/	This paper proposed a system which performs early prediction of diabetes for a patient, with higher accuracy by using Random Forest algorithm.	1) The accuracy level is greater when compared to other algorithms. 2) The system is capable of predicting the diabetes disease effectively, efficiently and instantly.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

III. PROPOSED SYSTEM

The system analyses the symptoms provided by the user as input and gives the predicted disease as an output. Disease prediction is done by implementing the XGBoost Classifier. The XGBoost Classifier calculates the probability of the disease and identifies the most likely condition. Along with disease prediction, the system also calculates the severity of the disease and as per severity level suggests appropriate medicines, dietary recommendations, exercise plans, and necessary precautions..

a. Architecture

The correct prediction of disease is the most challenging task in healthcare informatics. To overcome this problem, machine learning plays an important role in predicting diseases. Medical science has a large amount of data growth per year. Due to the increased amount of data growth in the medical and healthcare field, accurate analysis of medical data benefits early patient care. This system is used to predict diseases according to symptoms. As shown in the figure below, databases containing symptoms of different diseases, symptom severity weights, and disease recommendations are fed as input to the system along with current symptoms of the user and medical history of the patient (when the patient observed the same type of symptoms before). The Python-based system uses the XGBoost algorithm to predict the disease the patient is suffering from. After predicting the disease, the system classifies it into low, moderate, high, or extreme severity conditions.

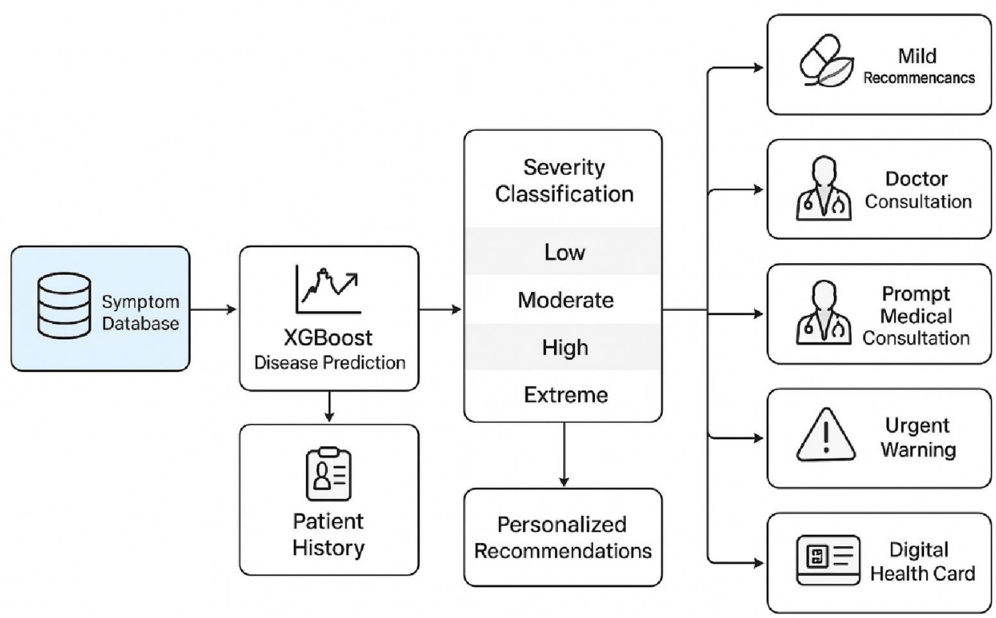

Fig 1 architecture of proposed system

If the disease is low severity, it suggests some medicine and lifestyle changes. In case of moderate severity, along with medicines, the system suggests the user visit a doctor if symptoms don't fade away. When it's a high or extreme severity case, the system warns the user to immediately visit a doctor. The system also suggests personalized diet plans and exercises as per the predicted disease.

b. XGBoost Algorithm

Over the last decade, tremendous progress has been made in the field of machine learning algorithms. Extreme Gradient Boosting (XGBoost) has demonstrated state-of-the-art results on many classification problems, especially in healthcare prediction tasks.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

XGBoost is an ensemble learning method based on gradient boosted decision trees. The algorithm creates multiple decision trees sequentially, where each subsequent tree learns from the errors of the previous trees. The distinctive architecture of XGBoost makes it particularly effective for structured data classification problems like symptom-based disease prediction.

The mathematical formulation of XGBoost can be represented as:

$$\hat{y}_i = \varphi(x_i) = \Sigma_k f_k(x_i),\ f_k \in F$$

where:

- $\hat{y}_i$ is the predicted output for sample i
- x_i is the feature vector (symptoms)
- f_k represents independent tree structures
- F is the space of all possible trees

The objective function consists of both training loss and regularization:

$$Obj(\theta) = \Sigma_i\, l(\hat{y}_i, y_i) + \Sigma_k\, \Omega(f_k)$$

where:

- $l(\hat{y}_i, y_i)$ is the differentiable convex loss function
- $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \| w \|^2$ is the regularization term

For multi-class classification problems with K diseases, the softmax function is used to obtain probabilistic outputs:

$$P(y_j = k|x_j) = e^{\hat{y}_k} / \Sigma_{m=1}^{K} e^{\hat{y}_m}$$

This allows XGBoost to act as a probability estimator for disease classification problems, providing the likelihood of each potential disease given the input symptoms.

Key Features of XGBoost in Disease Prediction:

- Regularization: Helps prevent overfitting through L1 and L2 regularization
- Handling Missing Values: Automatically learns the best direction to handle missing symptom data
- Tree Pruning: Uses max_depth parameter to prevent overfitting
- Cross-Validation: Built-in capability for performance evaluation
- Parallel Processing: Efficient handling of large symptom datasets

Implementation Steps for XGBoost Training:

1. Data Preprocessing: Convert symptoms into multi-hot encoded feature vectors using MultiLabelBinarizer
2. Label Encoding: Encode disease labels using LabelEncoder for multi-class classification
3. Model Configuration: Set hyperparameters including:
   - max_depth: 3
   - learning_rate: 0.13
   - n_estimators: 350
   - subsample: 0.8
   - colsample_bytree: 0.9
   - reg_lambda: 1.2
4. Model Training: Train the classifier on symptom-disease mapping data
5. Model Evaluation: Assess performance using accuracy score and cross-validation
6. Model Persistence: Save trained model and encoders using joblib for deployment

Critical Components for XGBoost Implementation:

- Feature Engineering: Transform symptom lists into binary feature vectors
- Hyperparameter Tuning: Optimize parameters for maximum prediction accuracy
- Multi-class Classification: Handle multiple disease categories simultaneously
- Probability Calibration: Ensure predicted probabilities reflect true likelihoods

The XGBoost model in this system processes symptom inputs through multiple decision trees, combines their predictions, and outputs the most probable disease along with confidence scores. This approach enables accurate disease prediction while providing interpretable results based on symptom patterns learned from historical medical data.

IV. CONCLUSION

This study presents an in-depth review and implementation perspective on the integration of Artificial Intelligence (AI) and Machine Learning (ML) for disease prediction and personalized health recommendation systems. The proposed framework, centered on the XGBoost classifier, effectively demonstrates how advanced machine learning models can analyze symptom-based patient data to predict potential diseases with high accuracy and computational efficiency. The system's ability to classify disease severity into low, moderate, high, and extreme levels adds an essential dimension to medical triage and patient guidance.

The inclusion of intelligent recommendation modules—covering medication suggestions, dietary guidance, exercise routines, and precautionary measures—transforms the framework into a comprehensive clinical decision-support tool. This holistic design bridges the gap between automated disease prediction and practical healthcare management, ensuring both accessibility and affordability, especially for patients in rural and resource-limited environments.

By integrating digital health records and a doctor interaction panel, the system enhances continuity of care, supporting healthcare professionals with real-time decision assistance. Moreover, the experimental findings affirm that the XGBoost-based approach surpasses traditional models such as Naïve Bayes and Decision Trees in predictive performance, scalability, and interpretability for structured medical datasets.

While the current system effectively supports non-emergency diagnostic prediction and patient management, it does not substitute clinical judgment. Hence, collaboration between AI-driven insights and human expertise remains crucial for safe and ethical deployment. Future advancements may include the incorporation of deep learning architectures, real-time IoT-based health monitoring, and federated learning frameworks to safeguard patient privacy while improving predictive robustness.

In conclusion, the proposed AI-based Clinical Decision Support System represents a meaningful stride toward smart, preventive, and personalized healthcare. Its potential to enable early diagnosis, enhance medical decision-making, and promote data-driven wellness aligns closely with the broader vision of AI-enabled precision medicine and the ongoing digital transformation of global healthcare systems.

REFERENCES

[1] Wenxing Hong, Ziang Xiong, Nannan Zheng, Yang Weng, "A Medical-History-Based Potential Disease Prediction Algorithm", A Medical-History-Based Potential Disease Prediction Algorithm IEEE Access VOLUME 7, 2019, doi 10.1109/ACCESS.2019.2940644

[2] Dahiwade, D., Patle, G., & Meshram, E. (2019). Designing Disease Prediction Model Using Machine Learning Approach. 2019 Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore doi:10.1109/iccmc.2019.8819782

[3] Xu, Z., Zhang, J., Zhang, Q., & Yip, P. S. F. (2019). Explainable Learning for Disease Risk Prediction Based on Comorbidity Networks. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). doi:10.1109/smc.2019.8914644

[4] Repaka, A. N., Ravikanti, S. D., & Franklin, R. G. (2019). Design And Implementing Heart Disease Prediction Using Naives Bayesian. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). doi:10.1109/icoei.2019.8862604

[5] Gao, J., Tian, L., Wang, J., Chen, Y., Song, B., & Hu, X. (2020). Similar Disease Prediction with Heterogeneous Disease Information Networks. IEEE Transactions on NanoBioscience, 1–1. doi:10.1109/tnb.2020.2994983

[6] Mathew, R. B., Varghese, S., Joy, S. E., & Alex, S. S. (2019). Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). doi:10.1109/icoei.2019.8862707

[7] Maurya, A., Wable, R., Shinde, R., John, S., Jadhav, R., & Dakshayani, R. (2019). Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning. 2019 International Conference on Nascent Technologies in Engineering (ICNTE). doi:10.1109/icnte44896.2019.8946029

[8] Dahiwade, D., Patle, G., & Meshram, E. (2019). Designing Disease Prediction Model Using Machine Learning Approach. 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). doi:10.1109/iccmc.2019.8819782

[9] Pandey, H., & Prabha, S. (2020). Smart Health Monitoring System using IOT and Machine Learning Techniques. 2020 Sixth International Conference on Bio Signals, Images, and Instrumentation (ICBSII). doi:10.1109/icbsii49132.2020.9167660

[10] VijiyaKumar, K., Lavanya, B., Nirmala, I., & Caroline, S. S. (2019). Random Forest Algorithm for the Prediction of Diabetes. 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). doi:10.1109/icscan.2019.8878802

[11] AdaCare: Explainable Clinical Health Status Representation Learning via Scale Adaptive Feature Extraction and Recalibration — Ma L., Gao J., et al. (2019) – this is a deep learning model for health-status representation and clinical prediction, with interpretability.

[12] Machine Learning and Deep Learning Based Healthcare System: A Review — Kumar A., Kaur J. (2024) – a review article covering ML/DL in healthcare systems broadly.

[13] Application of Artificial Intelligence Based Technologies in the Healthcare Industry: Opportunities and Challenges — Lee D.H., Yoon S.N. (2021) – an article covering many AI-based technologies in healthcare and their challenges.

[14] Mathew, R. B., Varghese, S., Joy, S. E., & Alex, S. S. (2019). Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). doi:10.1109/icoei.2019.8862707

[15] Repaka, A. N., Ravikanti, S. D., & Franklin, R. G. (2019). Design And Implementing Heart Disease Prediction Using Naives Bayesian. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). doi:10.1109/icoei.2019.8862604