International Journal of Emerging Research in Science, Engineering, and Management
Vol. 2, Issue 4, pp.103-108, April 2026.
www.ijersem.com eISSN- 3107-9075

Predictive Diagnostics for Chronic Disorders Using Machine Learning

[1]S. Savitha, [2]G. Jyothi Prakash, [3]M. D. Prasannalakshmi, [4]M. Karthik, [5]B. Manojkumar

Department of CSE, Siddharth Institute of Engineering & Technology, Puttur, AP, India.
[1]savithaselvaraj12@gmail.com, [2]mrjyothiprakash9199@gmail.com, [3]prassumd.2004@gmail.com, [4]karthikmandakala2004@gmail.com, [5]manojroyalz74@gmail.com

Abstract: In recent years, the integration of Artificial Intelligence into healthcare has opened new avenues for early diagnosis and preventive care. This study presents a comprehensive web-based platform designed to predict multiple diseases, namely diabetes, heart disease, Parkinson's disease, and breast cancer, using supervised machine learning algorithms. The system utilizes user-provided clinical parameters such as blood pressure, body mass index (BMI), glucose levels, age, and other relevant health indicators to generate accurate disease predictions. The backend prediction engine incorporates classification models, including Random Forest, Logistic Regression, and Support Vector Machine (SVM), trained and validated using publicly available medical datasets to ensure robustness and generalization. Unlike conventional single-disease prediction systems, the proposed framework provides multi-disease prediction within a unified interface, reducing the need for multiple diagnostic tools. The system also emphasizes interpretability by presenting prediction confidence scores and feature importance insights to improve transparency and reliability. The primary objective of the proposed framework is to support early disease detection and personalized healthcare through accessible predictive analytics, particularly in resource-limited regions. By combining machine learning-based predictive modeling with a user-friendly web interface, the system bridges the gap between computational intelligence and real-world clinical decision-making.

Keywords: Predictive Diagnostics, Chronic Disease Prediction, Random Forest, Logistic Regression, Healthcare Analytics.

1 INTRODUCTION

Chronic diseases such as diabetes, heart disease, Parkinson's disease, and breast cancer represent major global health challenges due to their long-term impact on mortality, quality of life, and healthcare systems. The prevalence of these conditions continues to increase because of lifestyle changes, aging populations, genetic predispositions, and environmental influences. Early detection and timely intervention play a crucial role in reducing disease severity, preventing complications, and improving patient outcomes. However, traditional diagnostic approaches often rely heavily on laboratory investigations, specialist consultations, and continuous clinical monitoring, which may not always be accessible in rural or resource-constrained regions. Recent advances in machine learning have enabled the development of intelligent predictive systems capable of identifying disease risks using structured clinical data such as age, blood pressure, glucose level, and body mass index (BMI). Supervised learning algorithms, including Random Forest, Logistic Regression, and Support Vector Machine (SVM), have demonstrated strong performance in detecting disease patterns and supporting early-stage diagnosis. These computational methods provide scalable and cost-effective alternatives to conventional diagnostic workflows by analyzing large volumes of patient health data efficiently and accurately.

Despite significant progress in predictive healthcare analytics, several challenges remain in deploying machine learning models in real-world medical environments. Clinical datasets often contain missing values, inconsistent measurements, and demographic biases that can affect prediction accuracy and reliability. Moreover, many existing prediction systems are designed to address only a single disease at a time and lack interpretability features that help clinicians understand prediction outcomes and confidence levels. To address these limitations, this work proposes a unified predictive diagnostics framework capable of estimating the risk of multiple chronic diseases within a single web-based platform. The proposed system integrates data preprocessing, feature extraction, and supervised machine learning classification techniques to generate accurate predictions along with interpretable confidence scores. By combining predictive analytics with an interactive user interface and contextual clinical insights, the system aims to support healthcare professionals and patients in making informed decisions and improving early disease detection, particularly in underserved regions with limited access to specialized diagnostic facilities.

2 LITERATURE REVIEW

Automated prediction of chronic diseases has evolved significantly from traditional statistical risk estimation techniques to advanced machine learning-based diagnostic frameworks. Early predictive systems primarily relied on demographic attributes and basic clinical parameters such as age, blood pressure, body mass index (BMI), and glucose levels. These features were commonly analyzed using classical statistical classifiers including Logistic Regression and k-Nearest Neighbors (k-NN), which provided moderate predictive performance in disease risk assessment tasks.

With the advancement of machine learning techniques, more robust classification algorithms such as Support Vector Machines (SVM), Random Forest, and ensemble learning models have been widely applied for chronic disease prediction. Among these approaches, Random Forest classifiers have demonstrated strong predictive capability due to their ability to handle nonlinear relationships, reduce overfitting, and process heterogeneous clinical datasets effectively. Comparative studies indicate that ensemble-based models frequently outperform individual classifiers when applied to diseases such as diabetes, heart disease, Parkinson's disease, and breast cancer.

Recent developments in deep learning have further enhanced predictive diagnostics by enabling automated feature extraction from structured clinical datasets. Deep neural networks and hybrid architectures have been successfully applied to capture complex relationships among clinical indicators, improving classification accuracy in multi-disease prediction tasks. These approaches reduce dependence on manual feature engineering and support scalable predictive modeling across diverse healthcare datasets. Transfer learning has also emerged as an effective strategy for improving disease prediction performance, particularly when labeled medical datasets are limited. By leveraging pretrained models and adapting them to healthcare-specific prediction tasks, researchers have achieved improved classification accuracy and reduced training complexity. This approach has shown promising results in applications such as chronic kidney disease prediction and cardiovascular risk assessment.

Despite these advancements, several practical challenges still limit the adoption of machine learning-based diagnostic systems in real-world healthcare environments. Many predictive models are trained using curated datasets that differ significantly from real clinical records, which often contain missing values, inconsistencies, and demographic imbalance. In addition, several existing systems focus primarily on improving classification accuracy without providing interpretable prediction outputs or confidence measures required for clinical decision support. Most currently available prediction platforms are designed for single-disease diagnosis and lack integrated frameworks capable of supporting multiple disease predictions simultaneously. Therefore, there remains a strong need for unified predictive diagnostic systems that combine robust machine learning models with interpretable outputs and user-friendly interfaces to improve accessibility, transparency, and reliability in healthcare decision-making.

3 PROPOSED METHOD

The proposed predictive diagnostics framework is designed to estimate the risk of multiple chronic diseases using structured clinical parameters and supervised machine learning techniques. The system integrates data preprocessing, feature extraction, classification modeling, and a user-friendly web interface to generate interpretable disease predictions along with confidence scores. The framework aims to support both healthcare practitioners and patients by providing accessible and reliable early diagnostic assistance.

3.1. System Overview

The overall workflow of the proposed system consists of four major stages: data collection, preprocessing, feature extraction, and disease prediction using machine learning classifiers. Clinical input parameters such as age, blood pressure, glucose level, and body mass index (BMI) are collected through a structured interface and processed before being passed to the prediction engine. The trained classification models analyze these features to estimate disease risk probabilities, which are presented to the user along with confidence measures and supporting clinical information. The overall workflow of the proposed predictive diagnostics framework is illustrated in Fig. 1.

3.2. Data Preprocessing

Data preprocessing plays a critical role in improving the reliability and performance of machine learning models. Clinical datasets often contain missing values, noise, and inconsistencies that can negatively affect prediction accuracy. Therefore, the collected data are first cleaned to remove incomplete or corrupted entries. Missing values are handled using appropriate imputation techniques to maintain dataset completeness. After cleaning, numerical features are normalized to ensure uniform scaling across all clinical parameters. Normalization prevents variables with larger numerical ranges from dominating the learning process and improves convergence performance in classifiers such as Support Vector Machines and Logistic Regression. The processed dataset is then prepared for feature extraction and classification.

3.3. Clinical Feature Representation

In the proposed system, structured clinical parameters are transformed into a standardized format suitable for machine learning classification. Relevant predictors such as blood pressure, glucose level, BMI, and age are selected as input features based on their clinical importance in chronic disease diagnosis. Feature normalization and encoding techniques are applied to ensure consistency and interpretability of the input data. Each processed patient record is treated as an independent feature vector representing the individual's clinical profile. These feature vectors are then used as input to supervised learning algorithms for disease risk prediction.

International Journal of Emerging Research in Science, Engineering, and Management
Vol. 2, Issue 4, pp.103-108, April 2026.
www.ijersem.com eISSN- 3107-9075

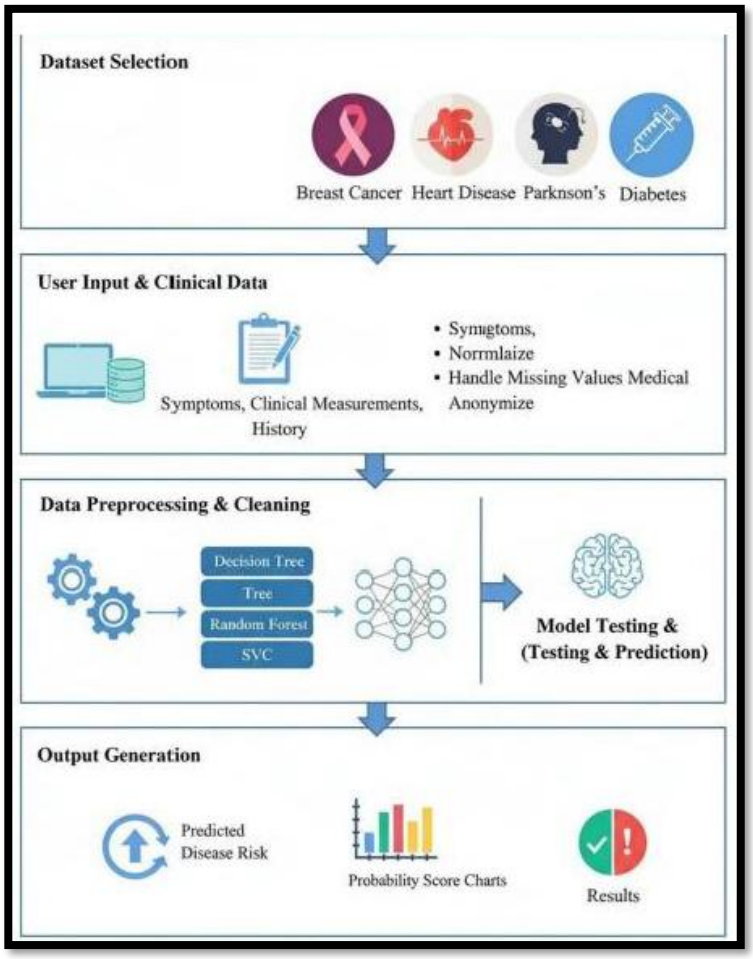

Fig. 1. System architecture of the proposed multi-disease predictive diagnostics framework.

3.4. Machine Learning Classification Model

The classification component of the proposed framework employs supervised machine learning algorithms including Random Forest, Logistic Regression, and Support Vector Machine classifiers. These algorithms are selected due to their proven effectiveness in healthcare prediction tasks and their ability to handle structured clinical datasets efficiently. During the training phase, labeled clinical datasets are used to train the classifiers to learn relationships between health parameters and disease outcomes. Model performance is optimized using validation datasets to ensure generalization capability and prevent overfitting. During inference, the trained models generate probability scores representing the likelihood of disease presence for each patient record.

3.5. Ensemble Prediction and Confidence Estimation

To improve prediction robustness and reliability, outputs from multiple classifiers are combined using an ensemble-based decision strategy. Ensemble averaging helps reduce prediction variance and improves overall classification stability compared to individual classifiers. The final prediction confidence score is computed by aggregating probability outputs from the classifiers. Let $p_i(c)$ represent the predicted probability of class $c$from classifier i, and let N be the number of classifiers used. The ensemble probability for class c is computed as:

$$P_{\text{ensemble}}(c) = \frac{1}{N}\sum_{i=1}^{N} p_i\,(c)$$

The class with the highest ensemble probability is selected as the final predicted disease category. This aggregated confidence score improves interpretability and helps users assess prediction reliability.

International Journal of Emerging Research in Science, Engineering, and Management
Vol. 2, Issue 4, pp.103-108, April 2026.
www.ijersem.com eISSN- 3107-9075

3.6. User Interface Integration

The proposed predictive diagnostics system is implemented through a web-based interface that enables users to input clinical parameters either manually or through structured data upload. After processing the input data, the system displays predicted disease risk along with confidence scores and ranked alternative risk probabilities. In addition to prediction results, the interface provides contextual clinical information such as disease description, symptoms, risk factors, and recommended follow-up actions retrieved from a structured knowledge base. This integration enhances usability and supports informed decision-making for both clinicians and patients.

4 Results and Discussion

4.1. Experimental Setup

The proposed predictive diagnostics framework was evaluated using publicly available clinical datasets corresponding to multiple chronic diseases, including diabetes, heart disease, Parkinson's disease, and breast cancer. The datasets were divided into training, validation, and testing subsets using stratified sampling to preserve class distribution across disease categories. Data preprocessing techniques such as normalization and missing value handling were applied prior to model training to ensure dataset consistency and improve prediction reliability. The classification performance of the proposed system was evaluated using widely accepted statistical metrics including Accuracy, Precision, Recall, and F1-score. These evaluation measures provide insight into prediction correctness, reliability of positive predictions, sensitivity toward actual disease cases, and the balance between precision and recall, respectively. These metrics are commonly used in medical decision-support system evaluation to validate prediction effectiveness. Fig. 2 shows the user interface designed to use the proposed system.

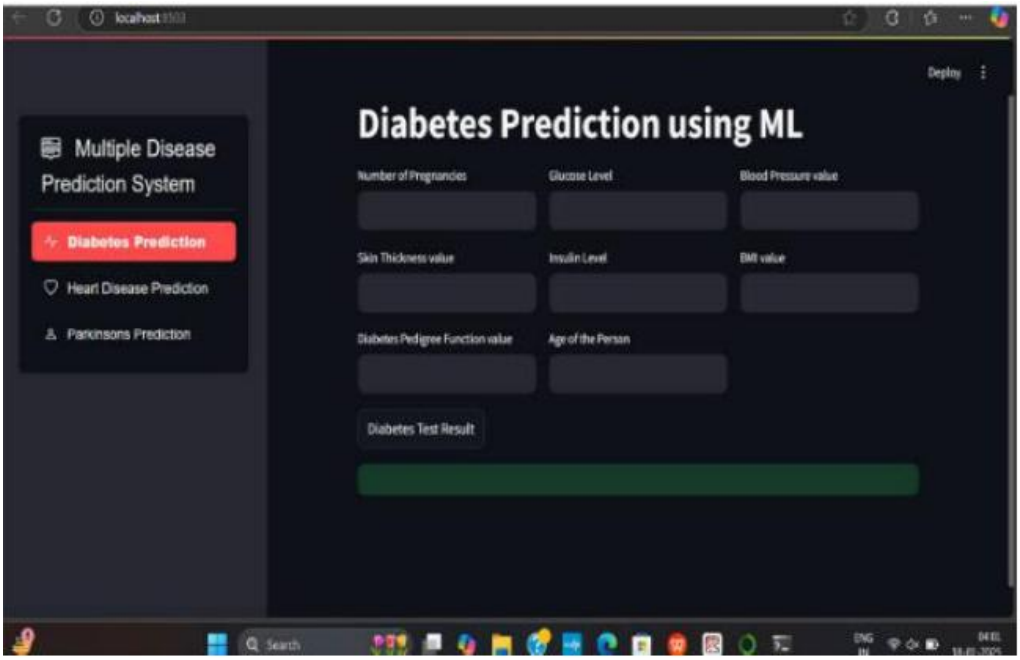

Fig. 2. User interface

4.2. Performance Evaluation Metrics

The evaluation metrics used for performance assessment are defined as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where TP represents true positives, FP represents false positives, and FN represents false negatives. These metrics ensure comprehensive evaluation of classification performance and support reliable comparison across different disease prediction models.

4.3. Overall Prediction Performance

The proposed multi-disease predictive diagnostics system demonstrated strong classification performance across the selected chronic disease datasets. Experimental results indicate that the model achieved an overall accuracy of 95.89%, with macro-averaged precision of 95.89%, recall of 95.84%, and F1-score of 95.84%. These results confirm that the proposed framework effectively distinguishes between multiple chronic disease conditions using structured clinical input parameters. The high performance of the model can be attributed to the integration of multiple supervised machine learning classifiers and the use of ensemble prediction strategies, which improve classification stability and reduce prediction variance across different disease categories.

4.4. Training and Validation Analysis

Training and validation accuracy curves indicate that the proposed classification framework successfully learns meaningful clinical feature relationships during early training stages and gradually stabilizes as convergence is achieved. The validation accuracy closely follows the training accuracy throughout the learning process, demonstrating that the model generalizes effectively to unseen clinical data. Similarly, validation loss follows a decreasing trend alongside training loss without significant divergence between the curves, indicating that overfitting is minimized through appropriate preprocessing and model optimization techniques. These observations confirm the robustness and reliability of the proposed predictive diagnostics framework.

4.5. Discussion

The integration of multiple supervised machine learning classifiers with ensemble-based prediction strategies contributes significantly to the strong performance achieved by the proposed system. Ensemble learning improves prediction stability by combining outputs from individual classifiers and reducing model variance compared to standalone classification approaches. Although the proposed framework demonstrates high classification accuracy across multiple chronic disease datasets, prediction errors may still occur in cases where clinical indicators overlap across disease categories. Incorporating additional contextual information such as patient lifestyle attributes, longitudinal health records, and demographic characteristics could further enhance prediction performance in future implementations. Furthermore, integrating explainable artificial intelligence techniques such as feature importance visualization and interpretable prediction scoring mechanisms can improve transparency and increase trust among healthcare practitioners using automated diagnostic support systems.

5 CONCLUSION

This paper presented a unified predictive diagnostics framework for early detection of multiple chronic diseases using supervised machine learning techniques. The proposed system integrates structured clinical data preprocessing, feature representation, classification modeling using Random Forest, Logistic Regression, and Support Vector Machine classifiers, and an ensemble-based prediction strategy to generate reliable disease risk estimates. In addition, the framework provides interpretable confidence scores through a user-friendly web-based interface, improving accessibility and usability for both healthcare practitioners and patients. Experimental evaluation conducted on publicly available chronic disease datasets demonstrated strong prediction performance, with overall accuracy, precision, recall, and F1-score exceeding 95%. These results confirm the effectiveness of the proposed system in identifying disease risk patterns across multiple chronic conditions using structured clinical indicators. Beyond quantitative performance improvements, the integration of ensemble confidence estimation and contextual disease information enhances prediction interpretability and supports informed clinical decision-making. The proposed framework therefore represents a practical and scalable solution for supporting early disease detection, especially in resource-constrained environments where access to specialized diagnostic infrastructure may be limited. Future work will focus on extending the system to include additional disease categories, incorporating longitudinal patient health records and lifestyle-related attributes, and integrating explainable artificial intelligence techniques to further improve prediction transparency and reliability in real-world healthcare applications.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ETHICS STATEMENT

This study did not involve human or animal subjects and, therefore, did not require ethical approval.

STATEMENT OF CONFLICT OF INTERESTS

The authors declare that they have no conflicts of interest related to this study.

REFERENCES

[1] N. H. Zainal and N. Van Doren, "Sleep disturbances predict nine-year panic disorder chronicity: The sleep-panic nexus theory with machine learning insights," *Journal of Anxiety Disorders*, vol. 114, p. 103052, Jul. 2025, doi: 10.1016/j.janxdis.2025.103052.S.

[2] Cheng *et al.*, "A machine learning approach to predicting postoperative recurrence in pediatric chronic rhinosinusitis: identification of key metabolic biomarkers," *American Journal of Otolaryngology*, vol. 46, no. 5, p. 104676, May 2025, doi: 10.1016/j.amjoto.2025.104676.

[3] L. Liang, T. Liu, W. Ollier, Y. Peng, Y. Lu, and C. Che, "Identifying new risk associations between chronic physical illness and mental health disorders in China: Machine Learning approach to a Retrospective Population analysis," *JMIR AI*, vol. 4, p. e72599, Apr. 2025, doi: 10.2196/72599.

[4] M. Olenik and H. M. Dönertaş, "Machine learning and OMIC data for prediction of health and chronic diseases," in *Elsevier eBooks*, 2025, pp. 365–388. doi: 10.1016/b978-0-323-95502-7.00284-0.

[5] K. Matsumura, K. Hamazaki, H. Kasamatsu, A. Tsuchida, and H. Inadera, "Decision tree learning for predicting chronic postpartum depression in the Japan Environment and Children's Study," *Journal of Affective Disorders*, vol. 369, pp. 643–652, Oct. 2024, doi: 10.1016/j.jad.2024.10.034.

[6] N. Almusallam and S. Khan, "Chronic liver disease classification using deep learning with SHAP-optimized hybrid features," *iScience*, vol. 28, no. 12, p. 113972, Nov. 2025, doi: 10.1016/j.isci.2025.113972.

[7] F. Zmudzki, R. J. E. M. Smeets, J. S. Groenewegen, and E. Van Der Graaff, "Machine Learning Clinical decision support for interdisciplinary multimodal chronic musculoskeletal pain treatment: Prospective Pilot Study of patient assessment and Prognostic Profile validation," *JMIR Rehabilitation and Assistive Technologies*, vol. 12, p. e65890, Feb. 2025, doi: 10.2196/65890.

[8] M. A. Shahbazi, M. A. Al-Mamun, T. Brothers, and I. Ahmed, "A machine learning framework for identifying phenotypes in chronic kidney disease," *Healthcare Analytics*, vol. 8, p. 100425, Oct. 2025, doi: 10.1016/j.health.2025.100425.

[9] J. Yang *et al.*, "Machine learning-based risk prediction of mild cognitive impairment in patients with chronic heart failure: A model development and validation study," *Geriatric Nursing*, vol. 62, no. Pt A, pp. 145–156, Feb. 2025, doi: 10.1016/j.gerinurse.2025.01.022.

[10] J. M and A. N, "Conceptual metaphor quantum correlation and radial basis extreme learning for predicting chronic kidney disease," *Computers & Electrical Engineering*, vol. 122, p. 109933, Dec. 2024, doi: 10.1016/j.compeleceng.2024.109933.

[11] L. Mauvieux *et al.*, "Artificial intelligence–based flow cytometry for the diagnosis of B-cell chronic lymphoproliferative disorders," *Blood Advances*, vol. 9, no. 22, pp. 5880–5887, Nov. 2025, doi: 10.1182/bloodadvances.2025016424.

[12] D. S. Khafaga, N. Khodadadi, E. Khodadadi, A. A. Alhussan, M. M. Eid, and E.-S. M. El-Kenawy, "Enhanced early chronic kidney disease prediction using hybrid waterwheel plant algorithm for deep neural network optimization," *Scientific Reports*, vol. 15, no. 1, p. 42584, Nov. 2025, doi: 10.1038/s41598-025-26382-6.