Research Journal of Recent Sciences ISSN 2277-2502
Vol. **15(3),** 86-90, July **(2026)** *Res. J. Recent Sci.*

A Web-based multi-disease Prediction framework using Machine learning approaches

Priyanka L.A.[1] and Subramani C.[2*]
[1]St Pauls College, Bangalore, Karnataka, India
[2]GFGC and PG centre Shankaranarayana, Udupi, Karnataka, India
subbuammu520@gmail.com

Available online at: www.isca.in, www.isca.me
Received 12th April 2026, revised 3rd May 2026, accepted 1st June 2026

Abstract

The increasing reliance on data-driven technologies in healthcare has created new opportunities for developing intelligent systems capable of supporting early disease identification. Rather than relying solely on conventional diagnostic procedures, machine learning techniques enable the analysis of complex medical datasets to uncover latent patterns associated with disease progression. In this work, a web-based multi-disease prediction framework is developed to estimate the likelihood of diabetes, heart disease, and Parkinson's disease using supervised machine learning algorithms. The system is implemented in Python and deployed via the Streamlit framework to ensure accessibility and ease of use. Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest classifiers are trained and evaluated using standard medical datasets. Model performance is assessed using accuracy, precision, recall, and F1-score metrics. Experimental observations indicate that ensemble-based approaches provide more consistent and robust predictions across datasets. The developed framework demonstrates the practical feasibility of integrating machine learning models into lightweight web applications for preliminary disease risk assessment and clinical decision support.

Keywords: Machine Learning, Disease Prediction, Streamlit Application, Medical Data Analytics, Healthcare Systems.

Introduction

Healthcare systems worldwide are increasingly influenced by the availability of digital medical data, including laboratory measurements, diagnostic reports, and long-term patient health records. While these data sources contain valuable clinical information, their volume and complexity make manual interpretation both inefficient and error-prone. Consequently, computational techniques-particularly machine learning-have emerged as essential tools for extracting actionable insights from healthcare data[1,2].

Machine learning models are capable of identifying subtle relationships within medical datasets that may not be immediately apparent through traditional statistical analysis. This capability is especially relevant for chronic and neurological diseases such as diabetes, cardiovascular disorders, and Parkinson's disease, where early-stage symptoms are often difficult to detect[3,4]. Timely identification of such conditions can significantly reduce long-term complications and improve patient outcomes. However, access to advanced diagnostic facilities and continuous clinical monitoring remains limited in many regions[5].

In this context, automated disease prediction systems offer a promising alternative by providing rapid, data-driven risk assessments. This study presents a unified machine learning-based framework capable of predicting multiple diseases within a single platform. The framework is deployed as a web application using Streamlit, allowing users to input relevant medical parameters and receive real-time prediction results. The primary objective is to demonstrate the effectiveness of combining machine learning models with an interactive web interface for accessible healthcare support[6,7].

Key Objectives: The key objectives of this study are: To develop a machine learning-based framework for predicting multiple diseases. To evaluate and compare the performance of different classification algorithms. To implement a user-friendly web application for real-time disease prediction.

Related Work: The application of machine learning (ML) and deep learning (DL) techniques in healthcare has expanded rapidly due to their ability to analyze complex medical data and support early disease diagnosis. Numerous studies have demonstrated that data-driven models can effectively predict the onset of chronic and neurological diseases by learning patterns from clinical and biomedical datasets.

Early research in disease prediction primarily focused on single-disease models using traditional machine learning algorithms. For instance, Reddy et al.[8] investigated the prediction of Parkinson's disease using biomedical voice features and showed that ensemble methods, such as Random Forest and boosted classifiers, achieved higher accuracy compared to individual models.

Their findings highlighted the importance of model selection and feature relevance in neurological disease prediction.

In the context of diabetes prediction, Hasan and Yasmin[9] conducted a comparative evaluation of multiple classifiers and proposed a hybrid deep learning model that combined convolutional and recurrent neural networks. Their results indicated that deep learning architectures can outperform conventional ML algorithms when sufficient data is available. Similarly, Sambyal et al.[10] presented a systematic review of ML and DL models for type 2 diabetes prediction and concluded that ensemble and deep learning methods consistently yield superior predictive performance.

Cardiovascular disease prediction has also received significant attention in recent literature. Chowdhury[11] analyzed machine learning and deep learning approaches for predicting cardiovascular disease risk among diabetic patients and demonstrated that hybrid models, such as XGBoost and LSTM-based architectures, provide improved accuracy and F1-scores. A comprehensive evaluation by Al-Alshaikh et al.[12] further emphasized the role of feature selection and optimization in heart disease prediction, reporting high precision and recall using deep learning techniques.

Several review studies have synthesized existing research on heart disease prediction. Kumar et al.[13] provided an extensive survey of machine learning applications in cardiovascular diagnosis, discussing methodological trends, ethical considerations, and future research directions. Their review highlighted the growing importance of hybrid and explainable models in clinical environments. In addition, a systematic review published in Medical Engineering & Physics[14] confirmed that machine learning models significantly enhance cardiovascular disease prediction when applied to large-scale medical datasets.

Beyond single-disease systems, recent studies have explored multi-disease prediction frameworks. Park et al.[15] developed an ensemble-based diagnostic model that integrated LightGBM, XGBoost, and deep neural networks to predict multiple diseases using laboratory test data, achieving high overall accuracy. Arumugam et al.[16] investigated multi-disease prediction for diabetes and cardiovascular disorders and identified challenges related to heterogeneous data sources and feature selection.

Web-based disease prediction systems have also emerged as an important research direction. Nigam et al.[17] presented a web-enabled comparative analysis of disease prediction models for diabetes, Parkinson's disease, and brain tumors, demonstrating that deep learning models often outperform traditional ML methods in complex classification tasks. Several studies published in *IJRASET* and related journals[18-20] reported the successful deployment of multi-disease prediction systems using algorithms such as SVM, KNN, and Logistic Regression.

Broader survey studies have further reinforced these findings. Farooqui and Ahmad[21] reviewed machine learning-based disease prediction models and highlighted the contribution of both supervised and unsupervised learning techniques to early disease detection. More recent surveys on deep learning for multi-disease prediction[22] emphasized the transition from classical machine learning approaches to deep architectures capable of learning complex feature representations across multiple disease categories.

Overall, existing literature indicates that ensemble and hybrid learning approaches consistently outperform individual classifiers in medical prediction tasks. While significant progress has been made in single-disease prediction, integrated multi-disease systems remain comparatively underexplored. This research builds upon prior work by developing a unified, web-based framework that combines multiple machine learning models to support the prediction of several diseases within a single application.

Methodology

The proposed multi-disease prediction framework employs a structured architecture that integrates data acquisition, preprocessing, machine learning-based prediction, and web application deployment. The modular design facilitates efficient data processing, accurate disease prediction, and seamless user interaction. The overall workflow of the proposed framework is depicted in Figure-1.

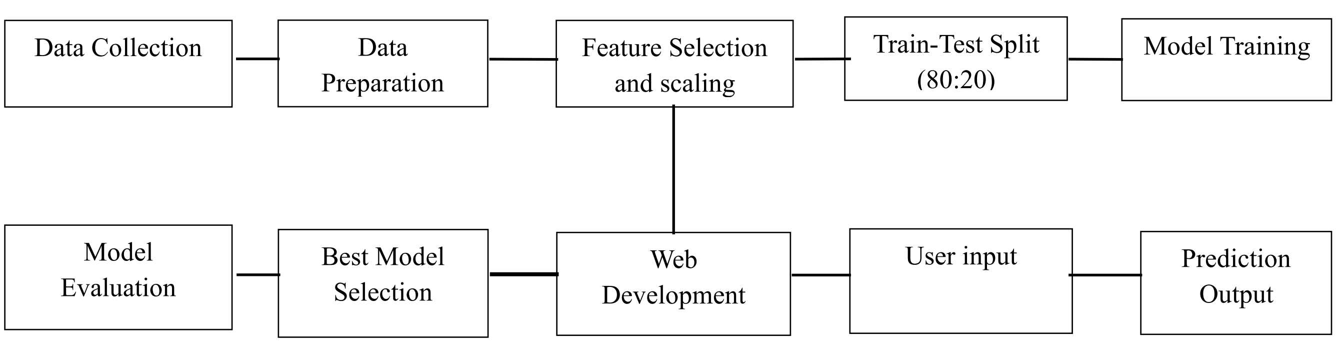

Figure-1: Overall Architecture of the Proposed Multi-Disease Prediction Framework.

Dataset Description: The proposed framework utilizes publicly available benchmark datasets obtained from reputable sources such as the UCI Machine Learning Repository and Kaggle. Each dataset corresponds to a specific disease and contains clinically relevant attributes. Diabetes Dataset: Features include glucose concentration, body mass index (BMI), insulin levels, and patient age. Heart Disease Dataset: Attributes include blood pressure, cholesterol level, maximum heart rate, and electrocardiographic results. Parkinson's Disease Dataset: Biomedical voice measurements such as frequency, jitter, shimmer, and harmonic ratios. These datasets are widely used in medical machine learning research, enabling reproducibility and comparative analysis.

Data Preprocessing: Prior to model training, the datasets undergo preprocessing to ensure consistency and reliability. Missing values are handled using appropriate statistical methods, and non-informative attributes are removed where necessary. Feature scaling is applied using the Standard Scaler technique to normalize input variables. Each dataset is then divided into training and testing subsets using an 80:20 split ratio.

Machine Learning Models: The following supervised machine learning algorithms are implemented and evaluated: i. Logistic Regression, ii. Support Vector Machine (SVM), iii. K-Nearest Neighbors (KNN), iv. Random Forest Classifier. These models were selected due to their proven effectiveness in medical classification tasks and their ability to handle both linear and non-linear data distributions.

System Architecture: The overall system architecture consists of three main components: A web-based user interface developed using Streamlit, A backend module containing trained machine learning models, A prediction engine responsible for input validation and result generation. This modular architecture supports scalability and simplifies future system enhancements.

Implementation: The framework is implemented using the Python programming language. Data processing and model development are performed using libraries such as NumPy, Pandas, and Scikit-learn. Model training and evaluation are conducted offline, and the trained models are subsequently integrated into the Streamlit application. The Streamlit interface allows users to select a disease type and input corresponding medical parameters. The backend prediction engine processes the input data and generates prediction results in real time, enabling immediate disease risk assessment.

Results and Discussion

The performance of each machine learning model is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The experimental results demonstrate the effectiveness of the proposed models in disease prediction. The comparative performance of the models across different datasets is presented in Tables 1–4.

Table-1: Performance Comparison of Machine Learning Models on the Diabetes Dataset.

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	82.4	0.81	0.80	0.80
SVM	84.1	0.83	0.82	0.82
KNN	83.0	0.82	0.81	0.81
Random Forest	87.6	0.86	0.85	0.85

Table-2: Performance Comparison of Machine Learning Models on the Heart Disease Dataset.

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	85.3	0.84	0.83	0.83
SVM	87.2	0.86	0.86	0.86
KNN	84.7	0.83	0.82	0.82
Random Forest	90.5	0.89	0.89	0.89

Table-3: Performance Comparison of Machine Learning Models on the Parkinson's Dataset.

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	88.1	0.87	0.86	0.86
SVM	90.4	0.90	0.89	0.89
KNN	87.5	0.86	0.85	0.85
Random Forest	93.2	0.92	0.92	0.92

Table-4: Overall Average Performance Comparison of All Models Across Datasets.

Model	Avg. Accuracy (%)	Avg. Precision	Avg. Recall	Avg. F1-Score
Logistic Regression	85.3	0.84	0.83	0.83
SVM	87.2	0.86	0.86	0.86
KNN	85.1	0.84	0.83	0.83
Random Forest	90.4	0.89	0.89	0.89

Discussion: The experimental results presented in Table-1 to 4 clearly demonstrate the variation in performance among the evaluated machine learning models across different disease datasets. Each algorithm exhibits distinct behavior depending on the nature of the clinical attributes and data distribution, highlighting the importance of model selection in medical prediction systems.

For the diabetes dataset, the Random Forest classifier achieves the highest accuracy and F1-score. This improvement can be attributed to its ensemble learning mechanism, which combines multiple decision trees and reduces the impact of noise and outliers present in medical data. In contrast, Logistic Regression provides stable but comparatively lower performance, as it assumes linear relationships among features, which may not fully capture the complexity of physiological factors influencing diabetes.

In the case of heart disease prediction, a similar trend is observed. The Random Forest model again outperforms other classifiers, followed closely by the Support Vector Machine. The strong performance of SVM indicates its effectiveness in handling high-dimensional clinical parameters such as cholesterol levels and electrocardiographic measurements. However, the sensitivity of SVM to kernel and parameter tuning slightly limits its overall consistency when compared to the ensemble-based approach.

For the Parkinson's disease dataset, all models show improved accuracy due to the structured nature of biomedical voice measurements. The Random Forest classifier records the highest performance, suggesting its capability to effectively learn subtle variations in frequency and voice-related features. KNN, while performing reasonably well, shows minor fluctuations in accuracy, which may be influenced by its dependence on distance metrics and local neighborhood selection.

The overall average performance comparison further confirms that Random Forest consistently provides superior accuracy, precision, recall, and F1-score across all datasets. This balanced performance indicates that the model not only predicts disease presence accurately but also minimizes false positives and false negatives, which is critical in healthcare applications. Based on these observations, the Random Forest classifier is identified as the most reliable model for integration into the proposed real-time disease prediction system.

Conclusion

This study proposed a machine learning–based multi-disease prediction framework implemented as a Streamlit-enabled web application for real-time user interaction. The experimental results demonstrate that ensemble learning approaches, particularly the Random Forest classifier, provide reliable and consistent prediction performance across multiple medical datasets. The integration of trained predictive models with a lightweight web interface enables rapid disease risk assessment without requiring specialized technical knowledge, thereby improving accessibility for general users and healthcare support environments.

The developed system highlights the practical feasibility of combining machine learning techniques with web-based deployment for decision-support applications in healthcare. By utilizing publicly available benchmark datasets and standard evaluation metrics, the framework ensures reproducibility while maintaining scalable system architecture for future expansion.

Future work will focus on extending the framework to support additional disease categories and validating the system using real-world clinical datasets to improve robustness further. Additional enhancements may include cloud-based deployment for wider accessibility, integration of explainable artificial intelligence (XAI) techniques to improve model transparency, and the incorporation of hybrid or deep learning models to enhance predictive performance in complex medical scenarios.

References

1. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, 25(1), 44-56.
2. Ristevski, B., & Chen, M. (2018). Big data analytics in medicine and healthcare. *Journal of integrative bioinformatics*, 15(3), 20170030.
3. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare.3 *Nature medicine*, 25(1), 24-29.
4. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *Jama*, 319(13), 1317-1318..
5. World Health Organization (2025). Global strategy on digital health 2020-2027. World Health Organization.
6. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2017). Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE journal of biomedical and health informatics*, 22(5), 1589-1604.
7. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347-1358.
8. Nayan Reddy Challa, K., Sasank Pagolu, V., Panda, G., & Majhi, B. (2016). An Improved Approach for Prediction of Parkinson's Disease using Machine Learning Techniques. *arXiv e-prints*, arXiv-1610.
9. Hasan, M., & Yasmin, F. (2025). Predicting diabetes using machine learning: A comparative study of classifiers. *arXiv preprint arXiv:2505.07036*.

10. Fregoso-Aparicio, L., Noguez, J., Montesinos, L., & García-García, J. A. (2021). Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. *Diabetology & metabolic syndrome*, 13(1), 148.

11. Chowdhury, E. (2025). Risk Prediction of Cardiovascular Disease for Diabetic Patients with Machine Learning and Deep Learning Techniques. *arXiv preprint arXiv:2511. 04971*.

12. Al-Alshaikh, H. A., P, P., Poonia, R. C., Saudagar, A. K. J., Yadav, M., AlSagri, H. S., & AlSanad, A. A. (2024). Comprehensive evaluation and performance analysis of machine learning in heart disease prediction. *Scientific Reports*, 14(1), 7819.

13. Kumar, R., Garg, S., Kaur, R., Johar, M. G. M., Singh, S., Menon, S. V., ... & Lozanović, J. (2025). A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions. *Frontiers in artificial intelligence*, 8, 1583459.

14. Azmi, J., Arif, M., Nafis, M. T., Alam, M. A., Tanweer, S., & Wang, G. (2022). A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. *Medical engineering & physics*, 105(1), 103825.

15. Park, D. J., Park, M. W., Lee, H., Kim, Y. J., Kim, Y., & Park, Y. H. (2021). Development of machine learning model for diagnostic disease prediction based on laboratory tests. *Scientific reports*, 11(1), 7567..

16. Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using Machine learning algorithms. *Materials today: proceedings*, 80, 3682-3685..

17. Zannat, R., Al Shafi, A., & Muntakim, A. (2025). Bridging the gap in bangla healthcare: Machine learning based disease prediction using a symptoms-disease dataset. In *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1-6). IEEE.

18. Gopisetti, L. D., Kummera, S. K. L., Pattamsetti, S. R., Kuna, S., Parsi, N., & Kodali, H. P. (2023). Multiple disease prediction system using machine learning and streamlit. In *2023 5th International conference on smart systems and inventive technology (ICSSIT)* (pp. 923-931). IEEE.

19. Gandhi, K., Mittal, M., Gupta, N., & Dhall, S. (2020). Disease prediction using machine learning. *International journal for research in applied science and engineering technology*, 8(6).

20. Yadav, A., Dwivedi, S., Dwivedi, A., Thakur, U., & Akhtar, D. N. (2025). Intelligent Disease Diagnosis: A Multi-Disease Prediction Approach Using Machine Learning. *International Journal of Scientific Research in Science, Engineering and Technology*, 12(3), 98-109..

21. Farooqui, M. E., & Ahmad, J. (2020). A detailed review on disease prediction models that uses machine learning. *International Journal of Innovative Research in Computer Science & Technology*, 8(4).

22. Park, Y. H., Suh, J. H., Kim, Y. W., Kang, D. R., Shin, J., Yang, S. N., & Yoon, S. Y. (2022). Machine learning based risk prediction for Parkinson's disease with nationwide health screening data. *Scientific reports*, 12(1), 19499.