# Smart Diagnosis: A Cross-Model Analysis for Predictive Healthcare Systems

***Prof.Dr. Vivek V. Kheradkar[1], Arya S. Hajare[2], Saniya H. Mulla[3], Ruturaj S. Kesare[4], Sanika D. Khane[5] ,Vidula C. Chopade[6] ,Girish R. Latkar[7]***

*1Assistant Professor,Department of Computer Science and Engineering, D.K.T.E's Soceity Textile & Engineering Institute, Ichalkarnji, Maharashtra, India.(vvkheradkar@gmail.com )*

*2Student,Department of Computer Science and Engineering, D.K.T.E's Soceity Textile & Engineering Institute, Ichalkarnji,Maharashtra, India.(aryahajareofficial@gmail.com )*

*3Student,Department of Computer Science and Engineering, D.K.T.E's Soceity Textile & Engineering Institute, Ichalkarnji,Maharashtra, India.(saniyamulla3036@gmail.com)*

*4Student,Department of Computer Science and Engineering, D.K.T.E's Soceity Textile & Engineering Institute, Ichalkarnji,Maharashtra, India.(ruturajkesare@gmail.com)*

*5Student,Department of Computer Science and Engineering, D.K.T.E's Soceity Textile & Engineering Institute, Ichalkarnji,Maharashtra, India.(sanikakhane@gmail.com)*

*6Student,Department of Computer Science and Engineering, D.K.T.E's Soceity Textile & Engineering Institute, Ichalkarnji,Maharashtra, India.(vidulachopade04@gmail.com)*

*7Student,Department of Computer Science and Engineering, D.K.T.E's Soceity Textile & Engineering Institute, Ichalkarnji,Maharashtra, India.(girishlatkar14@gmail.com)*

---

**Abstract -** In this study, we examine and contrast two approaches to disease prediction: a traditional machine learning-based system and a newly developed artificial intelligence-powered health monitoring model. The proposed framework leverages the Random Forest algorithm to enhance the reliability and precision of diagnosing critical health conditions, particularly heart attacks and strokes. Through a detailed technical evaluation, this paper underscores the advancements introduced by the new model, not only in terms of predictive accuracy but also in its overall system design.

A key feature of the enhanced model lies in its refined data preprocessing techniques, which allow for more accurate interpretation of patient data and reduction of noise that may affect prediction outcomes. Additionally, the system incorporates real-time alert mechanisms designed to notify individuals and healthcare providers when early warning signs are detected, enabling more proactive medical responses. Architectural innovations in the model contribute to improved scalability and efficiency, making it suitable for integration into broader health informatics systems.

By systematically analyzing the performance metrics and operational workflows of both systems, this paper demonstrates how the proposed AI-driven approach offers meaningful improvements over existing solutions. The findings point toward a promising direction for future development in intelligent healthcare systems, where machine learning models can play a critical role in preventive care and timely intervention.

***Key Words*:** HealthCare, Classification, Preprocessing, Prediction, Symptoms, Metrics.

## 1.INTRODUCTION

In today's rapidly evolving healthcare landscape, early detection of life-threatening conditions plays a critical role in enhancing patient care and reducing the burden on medical infrastructure. Traditional health systems have primarily focused on treatment rather than prevention, often leading to delayed diagnoses and increased risks. This paper introduces a proactive approach through an AI-powered disease prediction framework that utilizes machine learning techniques to identify early signs of severe illnesses like heart attacks and strokes.

The research involves a comparative assessment of several machine learning algorithms to determine their effectiveness in predictive accuracy and model efficiency. Among them, the Random Forest algorithm emerges as the most robust, achieving an accuracy of 95%. Its strength lies in combining multiple decision trees, allowing it to manage complex, high-dimensional datasets with ease.

Other models were also tested to gauge their suitability. The Weighted K-Nearest Neighbours (KNN), with a 91% accuracy rate, improves prediction by assigning relevance based on proximity. Fine KNN, although more sensitive to intricate data patterns, showed 80% accuracy. Medium KNN, with 63% accuracy, was less effective due to its generalized approach. The Naive Bayes classifier, reaching 94%, demonstrated strong performance with large-scale, structured data due to its probabilistic framework. Meanwhile, the Decision Tree algorithm achieved 92%, offering clarity in prediction pathways and ease of interpretation.

The developed system is enhanced with real-time monitoring and a responsive alert mechanism that can notify healthcare providers during critical scenarios. Especially beneficial in underserved or remote regions, this innovation shows how integrating AI with health analytics can transform preventive medicine by making early detection more precise, accessible, and actionable.

## 2. LITERATURE REVIEW

Machine Learning (ML) has emerged as a promising approach in the healthcare domain for predicting diseases based on medical datasets and patient-reported symptoms. Various studies have explored the application, comparison, and integration of ML algorithms to enhance diagnostic accuracy, early

detection, and decision-making processes in medical practice.

ML models such as Decision Tree, Naïve Bayes, and Support Vector Machine (SVM) have been compared to determine their effectiveness in disease classification, showcasing the ability of ML techniques to assist doctors in making informed decisions by analyzing large datasets [1]. Classification of 41 diseases using 95 key symptoms demonstrated the usefulness of ML in extracting insights from healthcare data, leading to early diagnosis and improved patient care [2].

The use of real-life patient parameters and algorithms like Random Forest, SVM, and AdaBoost highlights how ML can provide accurate and timely treatment support by enhancing healthcare decision-making [3]. Patient symptom analysis using real-world data further supports the efficiency and reliability of ML frameworks in medical diagnostics [4]. Leveraging patient-reported symptoms, ML models can assist healthcare professionals in identifying potential diseases early and efficiently [5].

Integrating big data analytics with ML allows examination of vast medical datasets to detect patterns and correlations that improve diagnostic accuracy, reflecting the growing significance of data-driven insights in modern healthcare [6]. Comparative evaluations of ML models based on performance metrics such as accuracy, precision, and recall reveal their varied suitability for healthcare applications, guiding model selection for optimal outcomes [7].

Supervised learning models including Decision Trees, SVMs, and Neural Networks have been assessed for their predictive capabilities, computational efficiency, and suitability for disease diagnosis, emphasizing the potential of ML in clinical environments [8]. Real-time health prediction systems using Random Forest and a Tkinter-based GUI interface illustrate how ML tools can be deployed effectively in user-friendly diagnostic applications [9].

Combining multiple ML techniques, such as Decision Trees, SVM, and Neural Networks, further enhances predictive accuracy and provides a robust framework for disease diagnosis [10]. The use of ML models for predicting multiple diseases like diabetes and heart conditions under a unified interface reflects efforts to address the scarcity of healthcare infrastructure and enable early intervention [11].

Ensemble approaches combining Logistic Regression, SVM, and KNN show high effectiveness in multi-disease prediction, supporting timely interventions and highlighting the value of integrated ML systems [12]. Evaluating models like Random Forest, Decision Tree, and LightGBM over 41 diseases confirms their ability to support medical professionals through high-accuracy predictions and targeted therapies [13].

Feature engineering, particularly through KNN and Fuzzy K-NN models, is shown to significantly enhance the prediction capabilities of ML systems by optimizing data preparation and model accuracy [14]. Deep learning methods such as CNNs and LSTMs offer high-performance prediction of chronic diseases, indicating the potential of hybrid deep learning architectures to revolutionize diagnostic systems [15].

16] Conducted a comparative study on ML models to classify diseases based on patient behavior and habits, investigating correlations between patient-related factors and diseases like diabetes and heart disease.

[17] Evaluated ensemble ML approaches for disease prediction using multiple datasets, highlighting the effectiveness of ensemble methods in improving prediction accuracy. [18] Reviewed the role of explainable artificial intelligence in disease prediction, emphasizing the importance of model transparency and interpretability. [19] Provided insights from a systematic literature review on AI for diabetes prediction, assessing datasets and ML algorithms like CNN, SVM, and XGBoost.

[20] Offered a comprehensive review of ML applications in medical prognostics, discussing techniques like Random Forest for sepsis prediction and CNNs for cancer detection.

## 3. METHODOLOGY

The proposed system for multiple disease prediction using machine learning is composed of several interlinked modules that work together to enable accurate diagnosis based on user-reported symptoms. The architecture follows a modular and layered design to ensure scalability, maintainability, and efficiency. The key components of the system are described below:

### 3.1 Data Preprocessing

- **Training Data**: The system begins with a medical dataset containing symptom-disease mappings, clinical indicators, and historical patient records.
- **Data Transformation**: Raw data undergoes cleaning, normalization, and encoding. This includes handling missing values, removing duplicates, and converting categorical variables into machine-readable formats.
- **Feature Extraction**: Relevant features are extracted using statistical and correlation-based techniques. This reduces dimensionality and improves model performance by selecting the most significant symptom indicators.
- **Processed Data**: The resulting clean, structured dataset is stored and utilized as the basis for training machine learning models.

### 3.2 Machine Learning Algorithms

- **Model Selection**: Four algorithms—Naïve Bayes, Random Forest, Decision Tree, and K-Nearest Neighbours (KNN)—are selected based on their interpretability, classification accuracy, and suitability for symptom-based prediction tasks.

- **Training**: Each algorithm is trained separately using the processed dataset. Cross-validation is employed to avoid overfitting and to ensure generalization to unseen user inputs.

### 3.3 User Interface and Input Module

- **User Details**: The system accepts basic demographic information from the user to personalize predictions and improve the contextual relevance of outcomes.
- **Symptom Input**: The user enters symptoms through a graphical or web-based interface. The system supports multi-symptom input for robust diagnosis.

### 3.4 Disease Prediction Model

- **Model Inference**: The symptom inputs are processed and passed through the trained machine learning models. Each algorithm runs in parallel to generate its own prediction.
- **Multi-Model Evaluation**: The system allows real-time comparison of predictions across algorithms, enabling better interpretability and offering the end-user the ability to assess consensus or divergence among model outputs.

### 3.5 Prediction Output

- **Naïve Bayes Output**: Probabilistic classification using Bayes' Theorem to estimate the likelihood of diseases.
- **Random Forest Output**: An ensemble of decision trees provides a majority-vote-based classification with high robustness against overfitting.
- **Decision Tree Output**: A single decision tree offers a hierarchical and interpretable structure for diagnosis based on feature splits.
- **KNN Output**: Computes the Euclidean distance between input symptoms and training data to classify diseases based on majority voting among k-nearest neighbours.

Each algorithm's output is displayed to the user, allowing for cross-verification and enhanced trust in the prediction system in comparative manner.

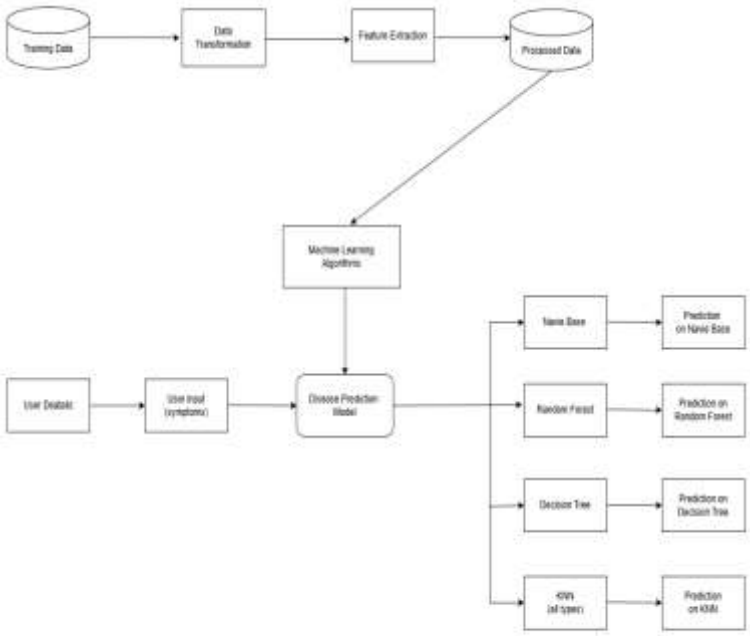


Fig 1.0 System Architecture

## 4. PERFORMANCE ANALYSIS

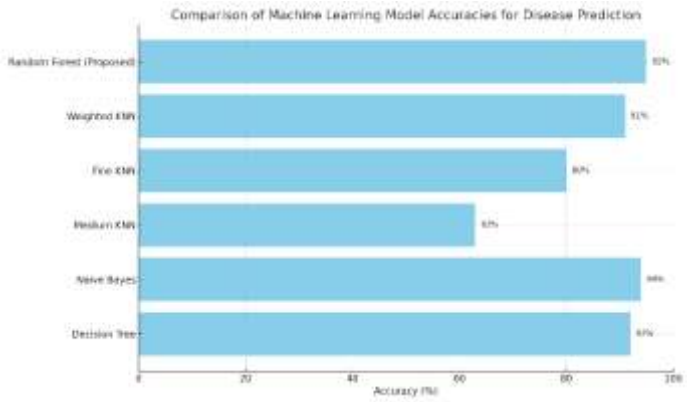


Fig 2.0 Comparative graph.

To highlight the distinction between existing techniques and the proposed approach, a visual comparison is presented in Comparative Graph. This graph illustrates a side-by-side evaluation of previous machine learning methods used for predicting diseases based on patient symptoms, showcasing how each performed in terms of accuracy.

The models included in this assessment span several widely used algorithms, such as Decision Tree, Naive Bayes, Medium KNN, Fine KNN, and Weighted KNN. Each of these techniques utilizes patient-reported symptoms to estimate the probability of disease presence. The graphical representation outlines the accuracy achieved by each method, drawing from earlier studies focused on symptom-based diagnosis.

Some models, particularly Naive Bayes and Weighted KNN, demonstrate strong performance with the added

benefit of low computational complexity. Naive Bayes is especially notable for its fast processing and suitability for large-scale data. Nonetheless, certain models—like Fine KNN and Medium KNN—face challenges related to data sensitivity and variance, which can hinder their consistency and generalizability.

The proposed approach, based on the Random Forest algorithm, addresses these limitations through its ensemble learning framework. By combining multiple decision trees and minimizing the risk of overfitting, it offers a more stable and accurate prediction process. The Random Forest model achieved a notable 95% accuracy, surpassing all other evaluated models in this comparison.

The graphical analysis confirms that while previous models like Naive Bayes and Weighted KNN reached accuracies of 94% and 91% respectively, and Decision Tree achieved 92%, the Random Forest approach leads with the highest precision.

Comparative Graph serves as a clear visual summary of these findings, emphasizing the strength of the proposed model in the context of early disease detection using symptom data.

TABLE I. Comparative Analysis

| Algorithm Used | Advantages | Limitations | Accuracy |
|---|---|---|---|
| Naive Bayes Classifier | Highly Scalable | Only for independent features & normally distributed data | 94% |
| Fine KNN | Improved accuracy | Sensitive to noise | 80% |
| Medium KNN | Simple, easy to implement | High Bias | 63% |
| Weighted KNN | Non-linear decision boundaries | Prone to overfitting | 91% |
| Random Forest (Proposed) | Ability to incorporate features & interactions | – | 95% |

## 5. CONCLUSION AND FUTURE SCOPE

This research presents a detailed evaluation of several machine learning algorithms—Random Forest, Decision Tree, Naive Bayes, and multiple K-Nearest Neighbour (KNN) variants—to determine their effectiveness in predicting diseases based on patient symptoms. Among the algorithms assessed, the Random Forest model delivered the highest accuracy at 95%, underscoring its effectiveness and consistency when applied to healthcare datasets.

Both Naive Bayes and Decision Tree algorithms also demonstrated strong potential, achieving accuracy rates of 94% and 92%, respectively, making them viable options for real-time implementation. Within the KNN family, Weighted KNN achieved better performance (91%) compared to Fine KNN (80%) and Medium KNN (63%), indicating that strategic parameter tuning and weighting significantly impact predictive success.

While each algorithm brings unique benefits, Random Forest distinguishes itself through its capacity to handle complex data structures and its resistance to overfitting—qualities essential for medical data analysis. These attributes make it a preferred choice for scalable and dependable disease prediction models.

Looking ahead, several enhancements can increase the model's practicality and adaptability in real-world healthcare environments:

**Wearable Device Integration** – Incorporating data from health monitoring devices (e.g., heart rate monitors, blood pressure trackers) could enable continuous tracking and more accurate predictions.

**Personalized Health Intelligence** – AI can be tailored to offer individual-specific insights and recommendations, such as preventive measures or lifestyle adjustments, to improve personal health outcomes.

**Inclusive User Interfaces** – Future versions of the system may feature voice-command input and multilingual support, ensuring broader accessibility for users in diverse and rural communities.

## 6. REFERENCES

1. C. K. Gomathy and A. R. Naidu, "The Prediction of Disease Using Machine Learning," *International Journal of Scientific Research* 2020.
2. K. Gaurav, A. Kumar, P. Singh, A. Kumari, M. Kasar, and T. Suryawanshi, "Human Disease Prediction Using Machine Learning Techniques and Real-Life Parameters," *International Journal of Advanced Research in Computer Science and Software Engineering*, 2021.
3. P. P. Reddy, D. M. Babu, H. Kumar, and S. Sharma, "Disease Prediction Using Machine Learning," *International Journal of Machine Learning Applications* 2021.
4. M. M. Ahsan, S. A. Luna, and Z. Siddique, "National Library of Medicine – National Center for Biotechnology Information," *Journal of Biomedical and Health Informatics* 2021
5. K. Gaurav, A. Kumar, P. Singh, A. Kumari, M. Kasar, and T. Suryawanshi, "Human Disease Prediction Using Machine Learning Techniques and Real-Life Parameters," *Journal of Machine Learning Research and Applications*, 2022.
6. M. Elhoseny, "Identification and Prediction of Chronic Diseases Using Machine Learning Approach," *IEEE Access* 2022.
7. A. Rathee, "Multiple Disease Prediction Using Machine Learning," *Journal of Artificial Intelligence in Medicine*, 2020.

8. B. Ramesh, G. Srinivas, P. R. P. Reddy, M. H. Rasool, D. Rawat, and M. Sundaray, "Feasible Prediction of Multiple Diseases Using Machine Learning," *Journal of Medical Systems*, 2023.
9. M. Venkatesh, "Multiple Disease Prediction Using Machine Learning, Deep Learning, and Streamlit," *Proceedings of the International Conference on Artificial Intelligence and Machine Learning in Healthcare*, 2023.
10. K. Saxena, R. Sharma, R. Kumar, and R. Kumar, "Disease Prediction Using Machine Learning and Deep Learning," *International Journal of Computer Applications in Health Sciences*, 2023.
11. "Multiple Disease Prediction Using Machine Learning Algorithms" by Chauhan et al. (2021):
12. "A Machine Learning Model for Early Prediction of Multiple Diseases to Cure Lives" by Kamboj et al. (2020)
13. "Symptoms Based Multiple Disease Prediction Model using Machine Learning Approach" by Kolli et al. (2021)
14. "Predictive Modeling for Multiple Diseases Using Machine Learning with Feature Engineering" by Krishnaiah et al. (2015)
15. "Multiple Disease Prediction Using Hybrid Deep Learning Architecture" by Al-Mallah et al. (2016)
16. Rinkal Keniya et al., 'Disease Prediction from Various Symptoms using Machine Learning', SSRN, 2020.
17. C. K. Gomathy and A. R. Naidu, 'The Prediction of Disease Using Machine Learning', Int. J. of Sci. Res., 2020.
18. J. A. Nahian, A. K. M. Masum, S. Abujar and M. J. Mia, "Common Human Diseases Prediction Using Machine Learning Based on Survey Data," *arXiv preprint arXiv:2209.10750*, 2022.
19. S. Mahapatra, R. Bandyopadhyay, P. Rathore, E. Elakiya and R. Sujithra, "Disease Prediction Using Machine Learning Techniques," *International Journal of Innovative Research in Science and Engineering (IJIRSE)*, vol. 2, no. 1, pp. 1–5, 2022
20. I. Gupta, V. Sharma, S. Kaur and A. K. Singh, "PCA-RF: An Efficient Parkinson's Disease Prediction Model Based on Random Forest Classification," *arXiv preprint arXiv:2203.11287*,2022

## BIOGRAPHIES

Dr. V. V. Kheradkar is P. hD in Computer Science & Engineering from Shivaji University, Kolhapur. Completed M.E. in Computer Science & Engineering, and an alumnus of WCE, Sangli where he has done his B.E. in computer science & engineering. He has 15 years of teaching experience and 3-month industrial experience. He has working as Assistant Professor in DKTE's, Textile and Engineering Institute, Ichalkarnji, Maharashtra, India. He has published his 6 research papers in national and international journals also four papers in national and international conferences. He has published 3 books. His has granted 3 patent and 6 patents are filled. His area of interest for research is uncertain and probabilistic databases and data structure.

Mr. Arya S. Hajare is a final-year B.Tech student at DKTE's Textile and Engineering Institute, Ichalkaranji, Maharashtra, India. His areas of interest include Cloud Computing and Web Technologies.

Miss. Saniya H. Mulla is a final-year B.Tech student at DKTE's Textile and Engineering Institute, Ichalkaranji, Maharashtra, India. Her areas of interest include DSA and Web Technologies.

Mr. Ruturaj S. Kesare is a final-year B.Tech student at DKTE's Textile and Engineering Institute, Ichalkaranji, Maharashtra, India. His areas of interest include System Design and Research.

Miss. Sanika D. Khane is a final-year B.Tech student at DKTE's Textile and Engineering Institute, Ichalkaranji, Maharashtra, India. Her areas of interest include DSA and Web Development.

Miss. Vidula C. Chopade is a final-year B.Tech student at DKTE's Textile and Engineering Institute, Ichalkaranji, Maharashtra, India. Her areas of interest include Database and Networking.

Mr. Girish R. Latkar is a final-year B.Tech student at DKTE's Textile and Engineering Institute, Ichalkaranji, Maharashtra, India. His areas of interest include JAVA Development and Machine Learning.