International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

Disease Prediction and Drug Recommendation Prototype using Multiple Machine Learning Algorithams

B. Hanumantha Rao[1], G Jayasri[2], Y Bindhu Sree[3], T Sandhya[4], A Chandhini Lakshmana Rai[5], P Maharshi[6]

[1]*Associate Professor, Department of AI and DS, PSCMR College of Engineering & Technology (Autonomous), Vijayawada – 520001, Andhra Pradesh, India*

[2,3,4,5,6]*Department of Computer Applications, PSCMR College of Engineering & Technology (Autonomous), Vijayawada – 520001, Andhra Pradesh, India*

***Abstract*-- The increasing demand for intelligent healthcare systems has accelerated the adoption of machine learning techniques for supporting early disease identification and clinical decision-making. This research presents a web-based Disease Prediction and Drug Recommendation System that predicts probable diseases from user-reported symptoms and provides appropriate medication recommendations together with disease information, precautionary measures, dietary guidance, and specialist consultation advice. The proposed framework employs multiple supervised machine learning algorithms, namely Random Forest, Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, and LightGBM, to perform comparative disease classification and identify the most effective predictive model. A comprehensive preprocessing pipeline consisting of data cleaning, feature encoding, and dataset transformation is implemented to improve data quality and model performance. The developed system is implemented using the Flask framework and integrates structured healthcare repositories for disease descriptions and drug recommendations to deliver comprehensive decision support through a user-friendly interface. Model performance is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score, to assess prediction reliability and generalization capability. Experimental analysis demonstrates that ensemble-based boosting methods provide superior predictive performance compared with conventional classification techniques, enabling more reliable symptom-based disease prediction. The proposed framework offers an efficient, scalable, and accessible healthcare assistance platform that supports preliminary diagnosis and informed healthcare decisions while reducing dependence on manual symptom assessment. The proposed approach can serve as a foundation for future intelligent healthcare applications through integration with electronic health records, explainable artificial intelligence techniques, and real-time clinical decision support systems.**

***Keywords*-- Disease Prediction, Drug Recommendation, Machine Learning, Light GBM, Random Forest, Flask, Healthcare Analytics, Clinical Decision Support, Symptom-Based Diagnosis.**

I. Introduction

Healthcare has witnessed significant advancements through the adoption of Artificial Intelligence (AI) and Machine Learning (ML), enabling intelligent systems that support disease diagnosis and clinical decision-making.

Early detection of diseases is essential for improving treatment outcomes and reducing healthcare costs. However, timely diagnosis is often limited by the availability of medical professionals, healthcare infrastructure, and diagnostic resources, particularly in rural and underserved regions. These challenges have increased the demand for automated healthcare systems capable of providing reliable preliminary disease assessment.

Machine learning techniques have demonstrated remarkable success in healthcare applications by identifying complex relationships between symptoms and diseases. Algorithms such as Random Forest (RF), Naïve Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and LightGBM have been widely employed for disease classification because of their ability to learn predictive patterns from clinical datasets. Nevertheless, the performance of these algorithms varies depending on the quality of the dataset and feature representation, making comparative evaluation essential for selecting an effective prediction model.

Although several symptom-based disease prediction systems have been reported, many are limited to disease classification and do not provide complete healthcare assistance. Most existing systems lack integrated drug recommendation, precautionary guidance, dietary suggestions, and specialist recommendations, thereby reducing their practical usefulness. Furthermore, many approaches rely on a single prediction model, which may affect prediction reliability across different medical conditions.

To address these limitations, this paper proposes a Disease Prediction and Drug Recommendation System Using Multiple Machine Learning Algorithms. The proposed framework employs Random Forest, Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, and LightGBM to predict diseases from user-reported symptoms. The system further provides medication recommendations, disease descriptions, precautionary measures, dietary advice, and specialist consultation information through a Flask-based web application. Comparative performance analysis is carried out using accuracy, precision, recall, and F1-score to identify the most suitable prediction model.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

The proposed system provides an intelligent and scalable healthcare decision-support platform that assists users in obtaining preliminary disease predictions and healthcare recommendations. The framework demonstrates the practical application of machine learning in improving healthcare accessibility and establishes a foundation for future enhancements involving explainable AI and real-time clinical decision-support systems.

II. Literature Review

Machine learning has become an important technology in modern healthcare by enabling intelligent systems capable of predicting diseases from patient symptoms. Various machine learning algorithms have been successfully applied to healthcare datasets to improve diagnostic accuracy and support clinical decision-making. These approaches assist healthcare professionals by providing faster and more reliable preliminary disease predictions.

Several researchers have developed disease prediction systems using supervised learning algorithms such as Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Random Forest improves prediction accuracy through ensemble learning. Naïve Bayes provides efficient probabilistic classification but assumes feature independence, which may reduce performance for complex medical datasets. Support Vector Machine performs well on high-dimensional data but requires careful parameter optimization. K-Nearest Neighbors classifies diseases based on similarity between patient records; however, its computational cost increases with larger datasets.

Recent studies have focused on advanced boosting algorithms, particularly LightGBM and XGBoost, because of their ability to process large healthcare datasets with improved speed and predictive performance. These algorithms effectively handle complex feature interactions and have demonstrated higher accuracy compared with several conventional machine learning models. Their efficiency makes them suitable for real-time healthcare prediction applications.

Although existing disease prediction systems have achieved encouraging results, many are limited to predicting diseases without providing complete healthcare support. Most applications do not include medication recommendations, precautionary measures, dietary guidance, or specialist consultation information. In addition, many studies rely on a single machine learning model, which may not provide consistent performance across different medical datasets.

To overcome these limitations, the proposed research develops a Disease Prediction and Drug Recommendation System Using Multiple Machine Learning Algorithms.

The framework integrates Random Forest, Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, and LightGBM for comparative disease prediction. In addition to identifying probable diseases, the system provides drug recommendations, disease descriptions, precautionary measures, dietary advice, and specialist guidance through a Flask-based web application. Comparative performance evaluation using Accuracy, Precision, Recall, and F1-score is performed to identify the most effective prediction model and to develop a reliable healthcare decision-support system.

III. Problem Statement And Objectives

A. Problem Statement

The increasing prevalence of diseases and the growing demand for timely medical diagnosis have highlighted the need for intelligent healthcare support systems. Conventional diagnostic methods depend heavily on clinical expertise, laboratory investigations, and hospital infrastructure, which may not always be accessible, particularly in rural and resource-limited areas. Although several machine learning-based disease prediction systems have been developed, many focus only on disease classification and provide limited post-diagnosis support. Most existing solutions do not integrate drug recommendations, precautionary measures, dietary guidance, and specialist consultation within a single platform. In addition, the use of a single prediction model may reduce the consistency and reliability of disease prediction across different medical datasets. Therefore, there is a need for an integrated healthcare system that combines multiple machine learning algorithms with comprehensive healthcare recommendations to improve preliminary diagnosis and support informed healthcare decisions.

B. Objectives

The primary objectives of the proposed research are:

1) To develop a symptom-based disease prediction system using multiple machine learning algorithms.
2) To preprocess and analyze healthcare data for improving prediction accuracy.
3) To compare the performance of Random Forest, Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, and LightGBM models.
4) To identify the most effective prediction model using Accuracy, Precision, Recall, and F1-score.
5) To provide appropriate drug recommendations, disease descriptions, precautionary measures, dietary suggestions, and specialist consultation guidance.
6) To implement the proposed framework as a user-friendly web application using the Flask framework.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

7) To develop an intelligent healthcare decision-support system that improves accessibility and supports preliminary medical diagnosis.

IV. Proposed Methodology

A. Dataset and Input Representation

The proposed Disease Prediction and Drug Recommendation System is designed to provide an intelligent healthcare assistance platform that predicts diseases based on user symptoms and recommends appropriate medications with healthcare guidance. The framework integrates data preprocessing, multiple supervised machine learning algorithms, and a Flask-based web application to deliver accurate and efficient disease prediction. Figure 1 presents the overall architecture of the proposed system.

A. Overall Framework

Initially, users select symptoms through the web interface. The collected symptoms are converted into a numerical feature vector and passed to the preprocessing module. The processed data are then supplied to six machine learning models, namely Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, and LightGBM. The predicted disease is forwarded to the recommendation module, which retrieves medicine information, precautionary measures, dietary suggestions, and specialist details from the healthcare repository before presenting the results to the user.

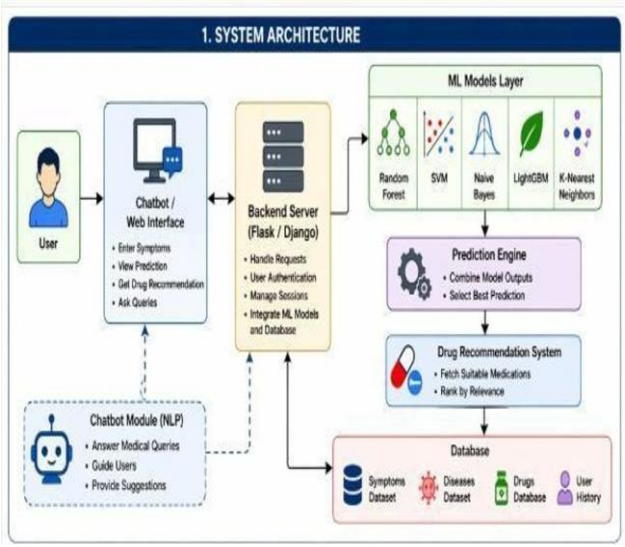


Fig. 1. Overall Architecture of the Proposed System

B. Data Preprocessing

The collected healthcare dataset is preprocessed to improve data quality before model training. Missing values and duplicate records are handled, categorical attributes are encoded into numerical values, and feature scaling is performed to normalize the input data.

The processed dataset is divided into training and testing subsets using an 80:20 ratio.

C. Model Development

The processed dataset is used to train six supervised learning algorithms, including Random Forest, Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, and LightGBM. Each model independently predicts the disease based on the user's symptoms. The prediction performance of all models is compared using Accuracy, Precision, Recall, and F1-score to identify the most reliable classifier.

D. Drug Recommendation Module

After disease prediction, the system retrieves disease-specific information from the healthcare repository. The recommendation module provides medicine names, dosage guidance, precautionary measures, dietary suggestions, and specialist recommendations through the Flask web application, enabling users to obtain preliminary healthcare guidance.

E. System Workflow

The workflow begins with symptom acquisition from the user, followed by preprocessing and feature transformation. The processed symptom vector is analyzed by multiple machine learning models to identify the most probable disease. The recommendation module then retrieves healthcare information and presents the final prediction results through the web interface.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

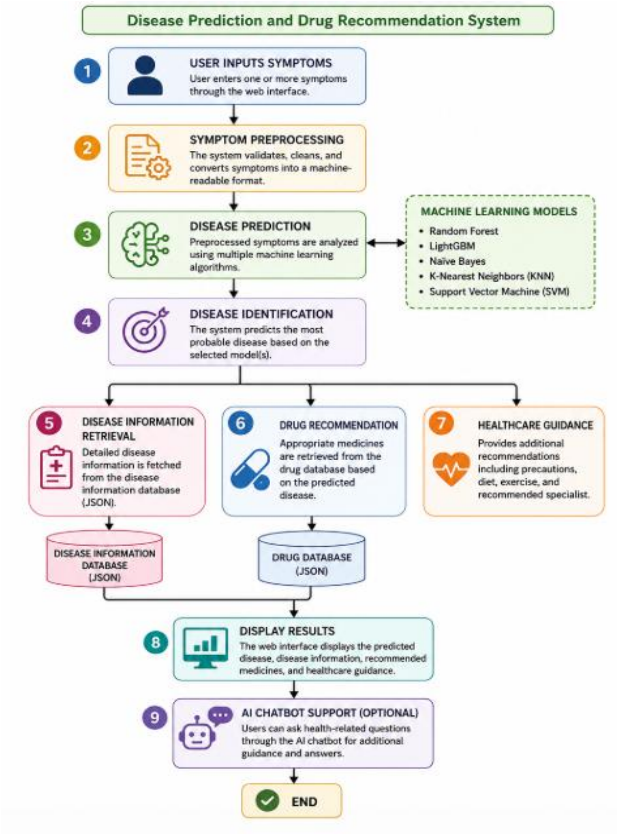


Fig. 2. Workflow of the Proposed System

F. Performance Evaluation

The effectiveness of the proposed framework is evaluated using Accuracy, Precision, Recall, and F1-score. These performance metrics provide a comprehensive assessment of the classification capability of each machine learning model and assist in identifying the most suitable model for disease prediction.

V. Results And Discussion

A. Experimental Results

The proposed Disease Prediction and Drug Recommendation System was evaluated using a symptom-based healthcare dataset through a Flask-based web application. To demonstrate the effectiveness of the framework, a sample test case was analyzed by selecting the symptoms **Shivering** and Continuous Sneezing. The processed symptom vector was evaluated by multiple machine learning classifiers, and the final prediction was generated using the integrated prediction framework.

TABLE I
Prediction Results for Sample Test Case

Parameter	Result
Selected Symptoms	Shivering, Continuous Sneezing
Predicted Disease	Allergy
Disease Severity	Mild
Recommended Specialist	Allergist
Final Prediction Confidence	**66.31%**

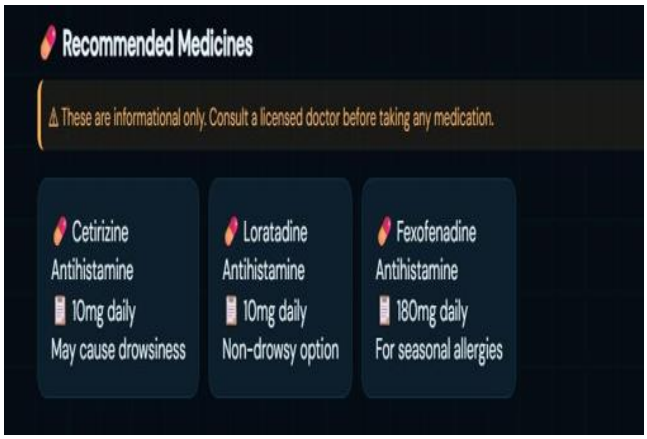
Table I shows that the proposed framework successfully identified **Allergy** as the most probable disease and provided the corresponding severity level and specialist recommendation.

TABLE II
Individual Model Prediction Confidence

Machine Learning Model	Predicted Disease	Confidence (%)
LightGBM	Allergy	77.00
Random Forest	Allergy	29.40
Naïve Bayes	Allergy	**100.00**
K-Nearest Neighbors	Allergy	**100.00**
Support Vector Machine	Allergy	25.20
Final Integrated Prediction	Allergy	**66.31**

The comparative analysis indicates that all machine learning models predicted the same disease, **Allergy**, although their confidence levels varied considerably. Naïve Bayes and K-Nearest Neighbors achieved the highest confidence for this sample, while Light GBM also produced a reliable prediction. Random Forest and Support Vector Machine predicted the same disease with lower confidence values. The integrated prediction framework combined the outputs of all models and generated a final confidence score of **66.31%**, providing a balanced and reliable diagnosis.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)



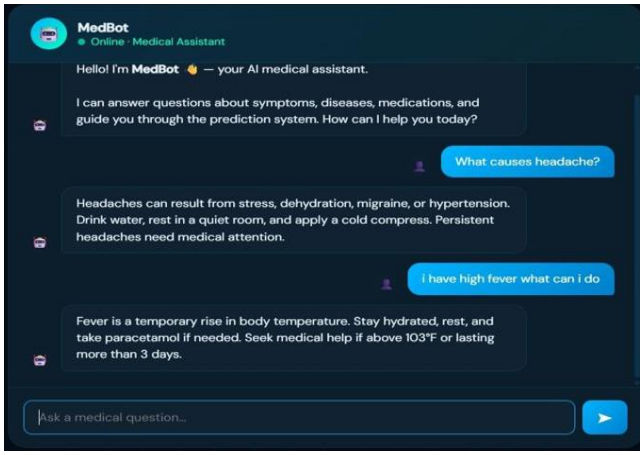


Fig. 3. Disease Prediction Result

The developed web application not only predicts the probable disease but also displays its severity, recommends an appropriate medical specialist, and provides additional healthcare guidance. This integration improves the practical usability of the proposed framework by supporting users with preliminary healthcare information before consulting a medical professional.

B. Discussion

The experimental results demonstrate that integrating multiple machine learning models improves the robustness of disease prediction by reducing dependence on a single classifier. Although individual models produced different confidence levels, they consistently identified the same disease for the selected symptoms. This consistency indicates that the proposed framework effectively captures symptom–disease relationships and provides reliable preliminary healthcare recommendations. The Flask-based implementation further enhances accessibility by delivering disease prediction and healthcare guidance through a user-friendly web interface.

VI. Conclusion

This paper presented a Machine Learning-Based Disease Prediction and Drug Recommendation System for supporting preliminary healthcare diagnosis using symptom-based analysis. The proposed framework integrates multiple supervised machine learning algorithms, including Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, and LightGBM, to predict diseases from user-provided symptoms. A Flask-based web application was developed to provide an interactive interface for symptom selection, disease prediction, and healthcare recommendations.

Comparative analysis demonstrated the effectiveness of employing multiple machine learning models for disease prediction. The integrated framework successfully identified the probable disease and provided additional healthcare support, including recommended medicines, precautionary measures, dietary guidance, and specialist consultation information. The implementation of the drug recommendation module enhances the practical applicability of the proposed system by delivering comprehensive healthcare assistance rather than limiting the output to disease prediction alone.

Overall, the proposed framework provides an intelligent, accessible, and scalable healthcare decision-support system capable of assisting users in obtaining preliminary medical guidance. The integration of machine learning with web technologies demonstrates the potential of intelligent healthcare applications to improve disease awareness and support early diagnosis while complementing professional medical consultation.

VII. Future Work

Although the proposed system provides reliable symptom-based disease prediction and healthcare recommendations, several enhancements can further improve its performance and applicability. Future work will focus on integrating deep learning techniques and explainable artificial intelligence (XAI) to improve prediction accuracy and model interpretability. The system can also be extended by incorporating electronic health records (EHRs), real-time patient monitoring through wearable devices, and cloud-based deployment to support remote healthcare services. Furthermore, the integration of large-scale clinical datasets and multilingual support can improve the scalability and accessibility of the framework, making it suitable for real-world intelligent healthcare applications.

REFERENCES

[1] S. K. Nayak, M. Garanayak, S. K. Swain, S. K. Panda, and D. Godavarthi, "An Intelligent Disease Prediction and Drug Recommendation Prototype by Using Multiple Approaches of Machine Learning Algorithms," *IEEE Access*, vol. 11, pp. 99304–99318, 2023, doi: 10.1109/ACCESS.2023.3314332.

[2] S. Dass, K. Jaiswal, and S. Kumar, "A Machine Learning-Based Recommendation System for Disease Prediction," SSRN Electronic Journal, 2025.

[3] J. Zhang, X. Zang, H. Chen, X. Yan, and B. Tang, "Revisiting Drug Recommendation From a Causal Perspective," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 2, pp. 1525–1533, 2025.

[4] X.-H. Yang, H.-J. Liao, P.-Y. Sun, J. Ma, B. Wang, Y. He, L.-G. Xue, L.-M. Su, and B.-J. Wang, "MCD-LightGBM System for Intelligent Analyzing Heterogeneous Clinical Drug Therapeutic Effects," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 6, pp. 3894–3905, 2025.

[5] N. Liu, Y. Tang, H. Yuan, L. Jiang, Z. Li, W. Zhang, and J. Wang, "Incomplete Multi-View Drug Recommendation via Multi-Level Representation Learning and Curriculum Learning," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2025.

[6] S. Maherunnisa and V. Walia, "Survey on Intelligent Disease Prediction and Drug Recommendation Using Machine Learning Algorithms," *Journal of Cardiovascular Disease Research*, vol. 16, no. 6, 2025.

[7] S. Naveed and M. Husnain, "A Drug Recommendation System Based on Response Prediction: Integrating Gene Expression and K-mer Fragmentation of Drug SMILES Using LightGBM," *Intelligence-Based Medicine*, vol. 11, Art. no. 100206, 2025.

[8] J. Omana, P. N. Jeipratha, K. Devi, S. Benila, and K. Revathi, "Personalized Drug Recommendation System Using Wasserstein Auto-Encoders and Adverse Drug Reaction Detection with Weighted Feed Forward Neural Network in Healthcare," *Journal of Informatics and Web Engineering*, vol. 4, no. 1, pp. 332–347, 2025.

[9] "Symptoms-Based Disease Prediction and Drug Recommendation Using Hybrid Machine Learning Algorithms," *IEEE Conference Publication*, 2024.

[10] J. Amutharaj, L. Gowda, M. Taufique, and K. M. Nanjunda, "Multiple Disease Prediction and Drug Recommendation System Using Machine Learning," *International Journal of Advanced Research in Innovative Ideas in Education (IJARIIE)*, vol. 11, no. 6, 2025.

[11] E. Y. Alqaissi, F. S. Alotaibi, and M. S. Ramzan, "Modem Machine-Learning Predictive Models for Diagnosing Infectious Diseases," 2022.

[12] J. Miah, D. M. Ca, M. A. Sayed, E. R. Lipu, F. Mahmud, and S. M. Y. Arafat, "Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction," 2023.