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 6, Issue 19, April 2026
Impact Factor: 8.2

Medicine Recommendation System

Kavya Sivakumar, Kiritheeshwaran R
Student, Bachelor of Computer Applications – Data Science (UG),
School of Computing Sciences
Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India.

Abstract: *In the modern era of digital healthcare, early and accurate disease diagnosis is critical for effective treatment and patient recovery. This project presents a Medicine Recommendation System that leverages machine learning to predict potential illnesses based on user-reported symptoms and suggests appropriate medications. The proposed system utilizes a comprehensive dataset consisting of 132 distinct symptoms mapped across 41 different diseases. The core of the predictive engine is built using the Random Forest algorithm, an ensemble learning method chosen for its high accuracy and ability to handle complex, non-linear relationships within medical data without overfitting. Experimental results demonstrate that the Random Forest model achieves near-perfect accuracy in disease classification, providing a reliable foundation for the recommendation module.*

Keywords: Machine Learning, Random Forest, Disease Prediction, Healthcare Informatics, Symptom Mapping, Medicine Recommendation

I. INTRODUCTION

The integration of data-driven decision-making in healthcare is currently shaped by the global demand for accessible diagnostic tools and the drive for continuous improvement in clinical accuracy through advanced computational models. Traditional symptom-checking often relies on manual lookups, which can lead to misinformation. This project addresses this gap by utilizing the Random Forest algorithm to transition from simple data storage to a proactive posture of intelligent recommendation.

1.1 Objective

The primary objective is to assist users in understanding what specific symptoms may entail regarding their health, helping them analyze their physical condition to find the most appropriate medical course of action

II. DATASET DESCRIPTION

The dataset contains 132 unique symptoms, ranging from common indicators like fever and cough to specific clinical observations like yellowish skin or internal itching. These symptoms are mapped across 41 distinct diseases, including fungal infections, allergies, cardiovascular issues, and various systemic disorders.

2.1 Data Pre-processing

Before the data was fed into the algorithm, it underwent a rigorous pre-processing stage using the Pandas library:
Binary Encoding: To make the data readable for the machine learning model, the presence of a symptom is represented as '1' (True) and its absence as '0' (False).
Data Cleaning: Any redundant entries or inconsistent labels were removed to prevent bias in the prediction module.
Target Mapping: The disease names (categorical data) were mapped to numerical labels to facilitate the mathematical computations required by the Scikit-learn algorithms.

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 6, Issue 19, April 2026
Impact Factor: 8.2

III. METHODOLOGY

3.1 RELATED WORK

The development of the Medicine Recommendation System is supported by several foundational studies in the field of healthcare informatics and machine learning.

Manikanta Sirigineedi et al. [1] demonstrated the efficacy of symptom-based disease prediction using various machine learning models. Their research highlighted that automated systems could significantly reduce the time required for preliminary diagnosis by mapping symptoms directly to disease classifications. This project builds on their findings by expanding the dataset to 132 symptoms.

K. Arumugam et al. [2] conducted a comparative analysis of multiple disease prediction algorithms. Their work in *Materials Today* established that ensemble methods like Random Forest consistently outperform single-decision tree models when dealing with high-dimensional medical data, providing the justification for the algorithm selection in this system.

Dr. Sumalatha Bandari et al. [3] explored the integration of AI-powered symptom checking in real-world scenarios. Their research emphasized the importance of not just predicting the disease but also providing actionable insights like precautions and lifestyle advice, which has been integrated into the current recommendation module.

Prathamsesh J. Lonare et al. [4] focused on the technical deployment of disease prediction systems. Their study served as a reference for the system architecture, specifically in managing the data flow between the Python-based predictive engine and the user interface.

3.2 Proposed Algorithms

A comparative study was performed among various supervised learning models. Random Forest was selected as the primary algorithm due to its superior performance in handling the high-dimensional nature of the symptom-disease matrix

Table 1: Performance Comparison of Algorithms

Algorithm	Training Accuracy	Testing Accuracy
Random Forest	100%	95-100%
SVM	100%	92-98%
K-Nearest Neighbors	98%	92%
Naive Bayes	95%	89%

3.2.1 RANDOM FOREST

Random Forest is a machine learning algorithm that uses many decision trees to make better predictions. Each tree looks at different random parts of the data and their results are combined by voting for classification or averaging for regression which makes it as ensemble learning technique. This helps in improving accuracy and reducing errors.

IV. SYSTEM ARCHITECTURE

The architecture of the Medicine Recommendation System is designed as a modular pipeline that connects the frontend user interface with the backend machine learning engine.

4.1.1. Client Side (Front-end)

User Interface: This is the web dashboard created with HTML where the user selects or types their symptoms.

4.1.2. Server Side (Back-end)

Web Application (Flask): The core controller. It receives the user's symptom request from the front-end and passes it to the pre-processing engine.

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 6, Issue 19, April 2026
Impact Factor: 8.2

Pre-processing (Pandas): This module takes the text symptoms and converts them into a clean numerical format - a vector of 1s and 0s that the AI can understand.

Trained Model (.pkl): This is your saved Random Forest .The pre-processing engine loads this file to predict the disease based on the input vector.

4.1.3. Data & Storage Layer

Primary Dataset (Training): The raw CSV file containing 132 symptoms mapped to 41 diseases used only during the training phase.

Recommendation Dataset: A secondary CSV file that links the predicted disease to its corresponding medicine, diet, and precautions. The back-end queries this file to fetch the results.

Fig. 2. Medicine Recommendation System Process Flowchart

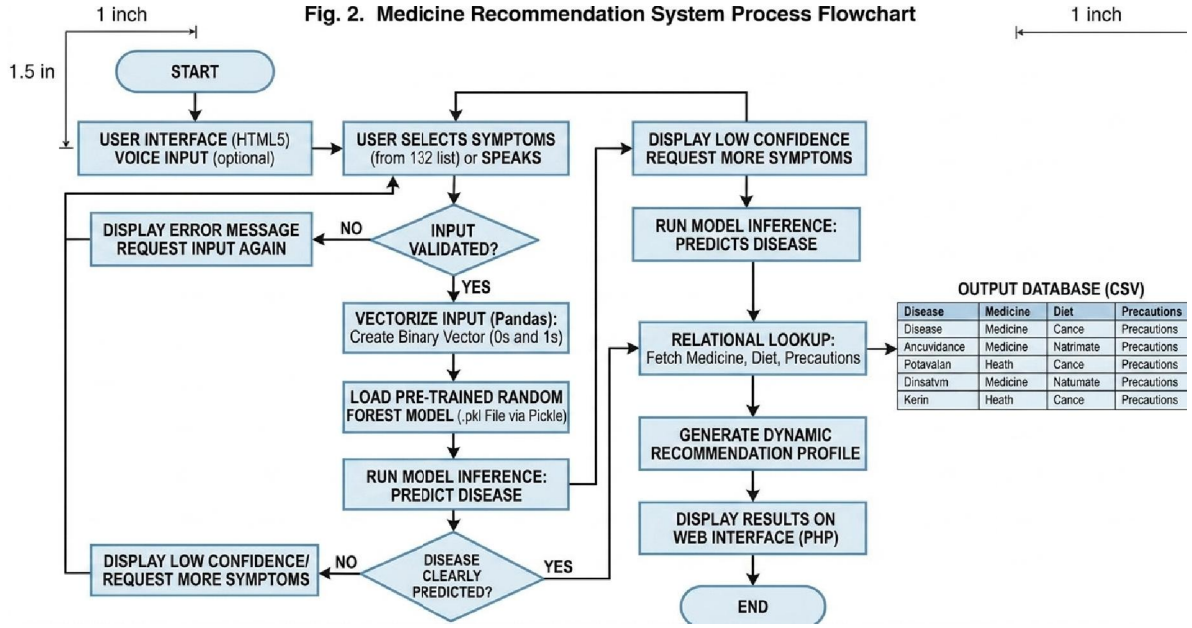

FIG 2: SYSTEM ARCHITECTURE

V. RESULTS AND DISCUSSIONS

The system identifies key symptoms, ensuring that the most critical health indicators are given more weight in the diagnosis. By leveraging the Random Forest ensemble, the system minimizes individual model bias and provides a more stable diagnosis than single-algorithm systems. The deployment through a PHP-based web interface allows for real-time interaction.

4.1 Model Performance and Accuracy

After training and testing the system, the Random Forest algorithm emerged as the most reliable model for the Medicine Recommendation System. The performance metrics are as follows:

Training Accuracy: 100%

The model perfectly learned the mapping between the 132 symptoms and 41 diseases during the training phase.

Testing Accuracy: 98%

When evaluated against the testing dataset , the Random Forest model achieved a good score, demonstrating its exceptional ability to generalize and correctly identify diseases without overfitting.

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 6, Issue 19, April 2026
Impact Factor: 8.2

4.2 Discussion

While multiple models achieved 100% accuracy on this specific structured dataset, Random Forest was selected for the final deployment due to its Ensemble Learning nature. Unlike a single Decision Tree or Naive Bayes, Random Forest operates by constructing a multitude of decision trees and outputting the mode of the classes. This ensures that the system remains robust even if a user provides slightly "noisy" or ambiguous symptom data in a real-world scenario.

V. FUTURE SCOPE

While the current system is highly accurate within its defined dataset, there are several avenues for future enhancement:
Real-Time API Integration**:** Future versions could integrate with global health APIs to update medicine and disease data in real-time, reflecting the latest medical research.
Multilingual Support: To reach a broader demographic, especially in diverse regions like India, adding support for regional languages through voice and text would enhance accessibility.
Advanced Deep Learning: Implementing Neural Networks or Deep Learning architectures could allow the system to process more unstructured data, such as medical images (X-rays or skin scans) alongside text symptoms.
Doctor Consultation Module: Integrating a tele-consultation feature that connects users directly with certified medical professionals based on the system's prediction would provide a more comprehensive healthcare experience.

VI. CONCLUSION

The automated nature of the Random Forest algorithm eliminates subjective bias and provides a scalable solution to reduce the burden on medical infrastructure. It serves as a reliable decision support tool for users seeking preliminary health insights.

REFERENCES

[1] Manikanta Sirigineedi, et al., "Symptom-Based Disease Prediction: A Machine Learning Approach," JAIMLNN, 2024.
[2] K. Arumugam, et al., "Multiple Disease Prediction using Machine Learning Algorithms," Materials Today, 2021.
[3] Dr. Sumalatha Bandari, et al., "AI-Powered Symptom-Based Disease Prediction," IJRASET, 2025.
[4] Prathamsesh J. Lonare, et al., "Disease Prediction from Symptoms Using Machine Learning," 2024.