Journal of Computer and Electrical Sciences (JCES)- Volume-1 Issue-4, April 2026
ISSN: 3049-2602

SYMPTOM BASED DISEASE PREDICTION SYSTEM FOR EARLY DETECTION OF DISEASES SUCH AS DENGUE, DIABETES, MALRIA

T. Bala Karthik[1],

Mrs.CH. Pavani[2]

[1]Student, Department of Computer Science & Engineering

Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India
[2]*Assistant Professor, Department of Computer Science & Engineering*
Andhra Loyola Institute of Engineering and Technology
Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India
Email id: thanugundlakarthik16@gmail.com

Abstract: *Disease prediction involves identifying the most probable disease based on observable symptoms using computational techniques. This project, titled "Symptom–Based Disease Prediction System for Early Detection," focuses on predicting diseases such as Dengue, Diabetes, and Malaria by analyzing user-provided symptoms. The system is trained on a structured dataset where symptoms are represented as binary features, enabling efficient pattern analysis across various diseases including Heart Attack, Pneumonia, Tuberculosis, Hepatitis, and more.Supervised machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes are used for prediction. Among these, Random Forest provides the highest accuracy and robustness, making it the final selected model. The trained model is saved using Joblib for efficient reuse.A user-friendly web interface is developed using Flask, HTML, and CSS, allowing users to select symptoms and receive real-time predictions along with probability, risk level, and explanations. The system serves as a decision-support tool to assist in early disease awareness and preliminary diagnosis.*

Keywords: Disease Prediction, Symptom-Based Diagnosis, Machine Learning, Random Forest, Healthcare Applications, Flask Web Application, Supervised Learning, Decision Support System.

1. INTRODUCTION

Recent advancements in healthcare data analysis and machine learning have enabled intelligent systems for disease prediction and early diagnosis. Early detection is essential for improving patient outcomes, but many individuals delay medical consultation due to lack of awareness or accessibility. Hence, automated systems can serve as effective decision-support tools.Disease prediction identifies the most probable disease based on observable symptoms. Machine learning models learn relationships between symptoms and diseases from medical datasets and can predict diseases using user-provided inputs.

The proposed "Symptom-Based Disease Prediction System" aims to predict diseases such as Dengue, Diabetes, and Malaria using a structured dataset containing multiple diseases including Pneumonia, Tuberculosis, Hepatitis, Typhoid, Hypertension, Allergy, and Migraine.Algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes are used, with Random Forest selected as the final model due to its higher accuracy. The model is stored using Joblib, and a web-based interface built with Flask, HTML, and CSS allows users to input symptoms and receive instant predictions with confidence levels and explanations.

2. Literature Survey

The development of machine learning–based disease prediction systems has gained significant attention due to the increasing need for early diagnosis and improved healthcare decision-making. Researchers have

explored various approaches using classification and deep learning techniques to analyze medical datasets and predict diseases based on symptoms. This section reviews existing approaches related to symptom-based disease prediction systems, focusing on accuracy, usability, and system integration.A study on disease prediction using machine learning algorithms such as Decision Tree, Random Forest, and Naive Bayes demonstrated that these models can effectively identify patterns in symptom data and assist in early detection. Similarly, research by Rajkomar et al. (2018) highlighted the effectiveness of deep learning models applied to electronic health records for improving diagnostic accuracy. Another study by Patel et al. proposed a web-based system that integrates machine learning models with a user-friendly interface for real-time disease prediction.

From the reviewed literature, the following key observations can be made:

- Machine learning algorithms can effectively predict diseases based on symptom patterns.
- Integration of predictive models with web applications improves accessibility and usability.
- Random Forest provides better accuracy and robustness compared to other algorithms.
- Real-time prediction systems enhance early awareness and support decision-making.
- Many existing systems lack detailed explanations and depend heavily on dataset quality.

3. Proposed System

To overcome the limitations of existing systems, this project proposes a Symptom-Based Disease Prediction System using Machine Learning. The system analyzes user-selected symptoms and predicts possible diseases using trained machine learning models. The proposed system utilizes a structured medical dataset containing symptoms and corresponding diseases. Machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes are used to learn patterns between symptoms and diseases. Among these algorithms, the Random Forest classifier is selected as the final model due to its higher accuracy and robustness. The trained model is integrated into a Flask-based web application, allowing users to easily select symptoms through a graphical interface. The system processes the selected symptoms, predicts possible diseases such as Dengue, Diabetes, Malaria, and other related 9 conditions, and displays results along with confidence percentages and explanations..

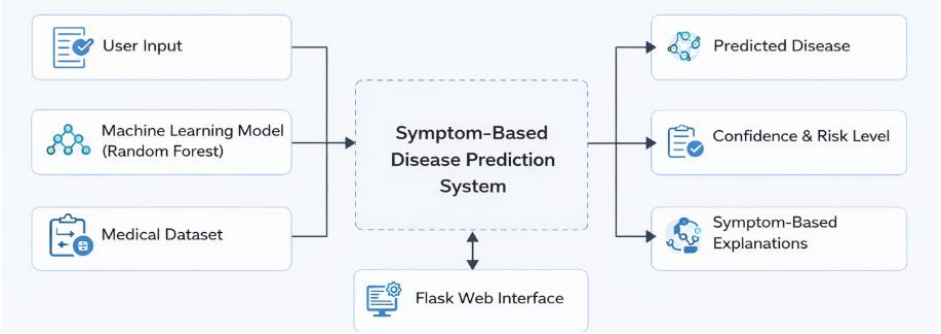

Fig 1: Proposed System

The application provides an interactive interface that allows users to view and control different functionalities such as:

- Provides automatic disease prediction based on symptoms.
- Uses machine learning algorithms to improve prediction accuracy.
- Offers a user-friendly web interface for easy interaction
- Displays confidence levels and risk indicators for predicted diseases.
- Provides detailed explanations including matching and missing symptoms.
- Helps in early disease awareness and preliminary diagnosis.

4. Methodology

The methodology of the proposed system is organized into the following steps:

1. User Input (Interface Module):The system provides a web-based interface where users can select symptoms using a checkbox-based design. The interface is developed using HTML, CSS, and Bootstrap to ensure simplicity and ease of use.
2. Data Transmission:Once the user selects symptoms, the input data is sent to the backend server for processing. The selected symptoms are converted into a structured format suitable for machine learning models.
3. Web Application Processing:The Flask-based web application acts as the central controller. It receives user inputs, processes the data, and forwards it to the machine learning prediction module.
4. Machine Learning Prediction:The system uses trained machine learning models such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes. Among these, the Random Forest model is selected due to its higher accuracy and robustness. The model predicts the most probable disease based on input symptoms.
5. Dataset Utilization:The system is trained on a structured medical dataset where symptoms are represented as binary values (present/absent). This dataset enables the model to learn relationships between symptoms and diseases.
6. Model Deployment:The trained model is stored using the Joblib library, allowing it to be loaded quickly for real-time predictions without retraining.
7. Result Generation:The system generates prediction results including the most probable disease along with confidence percentages and risk levels.
8. User Output Display:The predicted results, along with explanations such as matching and missing symptoms, are displayed to the user through the web interface, supporting early awareness and preliminary diagnosis.

5. Proposed System Results

The proposed Symptom-Based Disease Prediction System was successfully developed and tested under different usage scenarios. The system effectively predicts diseases based on user-selected symptoms and provides accurate results in real time.

• The system accurately processes user-input symptoms through a structured web interface. The symptom selection process was simple, responsive, and user-friendly.

• The machine learning model, particularly the Random Forest classifier, demonstrated high accuracy in predicting diseases such as Dengue, Diabetes, Malaria, and other related conditions.

• The system successfully generated predictions along with confidence percentages, helping users understand the likelihood of each disease.

• The application provided detailed explanations including matching and missing symptoms, improving transparency and user understanding of the results.

• The web application built using Flask performed efficiently, with quick response time and smooth interaction between frontend and backend.

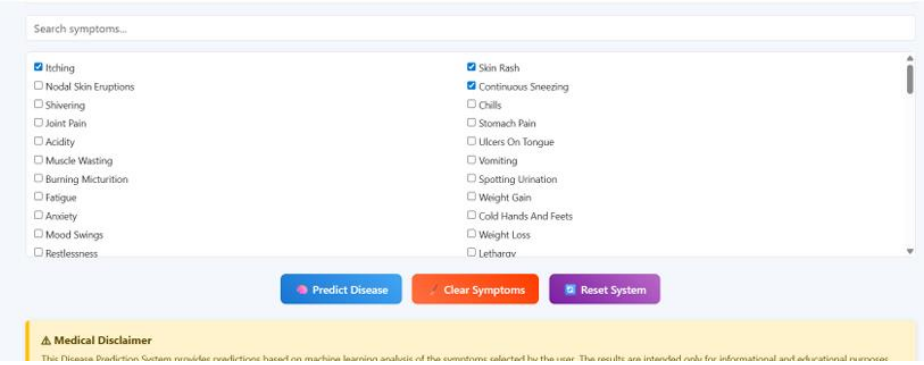

Fig 2: Symptom Selection Interface

The system response time was fast, and predictions were generated instantly without noticeable delay.

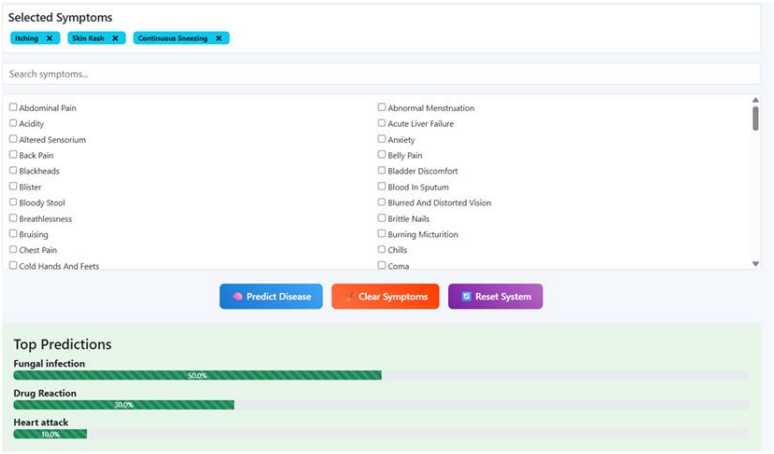

Fig 3: Prediction Output

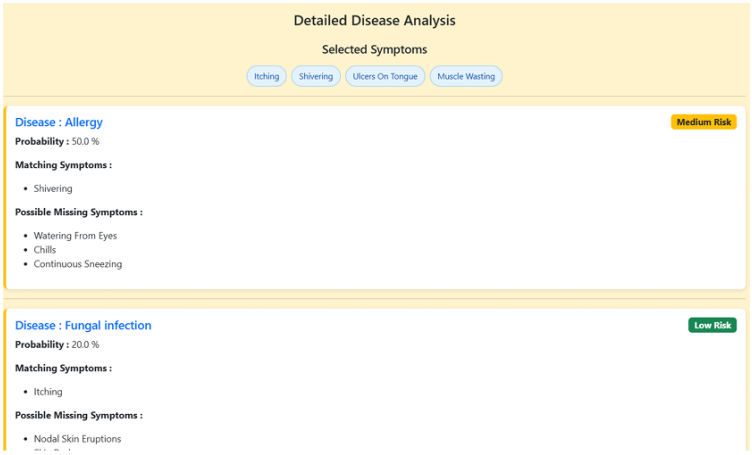

Fig 4: Explanation and Confidence Display

Overall, the system demonstrated reliable performance in disease prediction, real-time processing, and user interaction. It effectively supports early disease awareness and acts as a decision-support tool for preliminary diagnosis.

6. CONCLUSION

The Symptom-Based Disease Prediction System is developed to predict possible diseases based on symptoms provided by the user. The system uses machine learning techniques to analyze symptom patterns and identify the most probable disease. The application is implemented using Python, Flask, HTML, CSS, and JavaScript, which provide a simple web interface and efficient backend processing. Machine learning algorithms such as Decision Tree, Random Forest, Gradient Boosting, and Naive Bayes are used for training the model. Among these, the Random Forest algorithm is selected as the final prediction model due to its higher accuracy. The system allows users to select symptoms through an interactive web interface and receive disease predictions along with confidence levels. Features such as input validation, prediction display, and symptom reset options improve usability. Testing confirms that the system performs correctly and produces accurate predictions based on the trained dataset. Although the system does not replace professional medical diagnosis, it can serve as a support tool for early disease awareness and preliminary health analysis. The project demonstrates how machine learning and web technologies can be combined to develop intelligent healthcare applications.

REFERENCES

[1] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences, 2017. Available: https://archive.ics.uci.edu
[2] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," Informatica, vol. 31, no. 3, pp. 249–268, 2007.
[3] L. Breiman, "Random Forests," Machine Learning Journal, vol. 45, no. 1, pp. 5–32, 2001.
[4] A. Rajkomar, E. Oren, K. Chen et al., "Scalable and Accurate Deep Learning with Electronic Health Records," NPJ Digital Medicine, vol. 1, no. 18, 2018.
[5] T. Mitchell, Machine Learning, McGraw-Hill Education, 1997.
[6] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd Edition, Morgan Kaufmann Publishers, 2011.
[7] F. Chollet, Deep Learning with Python, Manning Publications, 2018.
[8] Kaggle Dataset, "Disease Prediction Using Machine Learning Dataset," Available: https://www.kaggle.com/datasets/kaushil268/disease-prediction-using-machine-learning