International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 03 | Mar 2026 www.irjet.net p-ISSN: 2395-0072

Smart Symptom Driven Disease Diagnosis and Prevention Recommendation System

Mrs.B.Priyanka[1], Bhargavi Guggilla[2], Sindhuja Gummula[3], Shiva Gilla[4]

[1] Associate Professor, Department of Computer Science and Engineering
[2,3,4] B.Tech Students, Department of Computer Science and Engineering
Teegala Krishna Reddy Engineering College , Telangana, India

Abstract - *Early identification of diseases based on symptoms plays a significant role in improving healthcare outcomes and preventing severe complications. However, traditional diagnosis methods often require medical consultation and may take time for accurate evaluation. To address this challenge, this research proposes a Smart Symptom-Driven Disease Diagnosis and Prevention Recommendation System using Machine Learning techniques. The system allows users to input symptoms through an interactive interface and predicts the most probable disease using trained machine learning algorithms such as Naive Bayes and Random Forest. The dataset used for training contains multiple symptoms and associated diseases, enabling the model to learn relationships between symptoms and medical conditions. In addition to predicting the disease, the system also provides preventive recommendations to guide users in managing their health condition at an early stage. Furthermore, an AI-powered chatbot is integrated into the system to assist users by providing medical information and guidance based on their queries. The experimental results demonstrate that the proposed system achieves reliable prediction accuracy and provides useful preventive suggestions. The developed solution can serve as a preliminary decision-support tool for users by offering quick insights into possible health conditions and recommended precautions. This approach contributes toward improving accessibility to basic healthcare information through intelligent technologies.*

Key Words: **Disease Prediction, Machine Learning, Naive Bayes, Random Forest, Healthcare AI, Symptom Analysis, Prevention Recommendation System**

1.INTRODUCTION

Healthcare systems around the world are increasingly adopting intelligent technologies to improve the efficiency and accuracy of disease diagnosis. Early detection of diseases based on symptoms can significantly reduce the risk of complications and improve treatment outcomes. However, traditional diagnostic procedures often require direct consultation with medical professionals, which may not always be immediately accessible to individuals. With the rapid advancement of machine learning and artificial intelligence technologies, it has become possible to develop intelligent systems that assist users in identifying potential diseases based on their symptoms.

Machine learning techniques have demonstrated significant potential in the medical domain for analyzing complex health data and predicting diseases with improved accuracy. According to Kononenko [2], machine learning models can assist medical practitioners by identifying patterns within healthcare datasets and providing reliable diagnostic predictions. Algorithms such as Naive Bayes and Random Forest are widely used for classification problems, including medical diagnosis, due to their efficiency and ability to handle structured datasets [5][6]. These algorithms can learn relationships between symptoms and diseases by analyzing historical medical datasets.

Recent studies have highlighted the growing role of artificial intelligence in healthcare applications. Rajkomar et al. [8] emphasize that machine learning can help improve diagnostic accuracy by analyzing patient data and providing predictive insights. Similarly, Deo [9] discusses how intelligent systems can assist in identifying diseases early by processing large amounts of medical data and generating meaningful predictions. By utilizing machine learning algorithms, automated diagnostic systems can support healthcare decision-making and enhance accessibility to medical information.

Despite these advancements, many existing healthcare systems still rely on manual symptom evaluation, which can be time-consuming and may not always provide immediate feedback to patients. To address this limitation, a symptom-driven disease prediction system can be developed to analyze user-reported symptoms and predict possible diseases using trained machine learning models. Such systems can act as preliminary diagnostic tools that assist users in understanding their health conditions before seeking professional medical consultation.

In addition to disease prediction, providing preventive recommendations is equally important for improving healthcare awareness. Preventive guidance allows users to take early precautions and adopt healthier habits to reduce the risk of disease progression. Therefore, integrating disease prediction with prevention recommendations can significantly enhance the usability and practical value of intelligent healthcare systems.

The proposed system, titled "Smart Symptom-Driven Disease Diagnosis and Prevention Recommendation System Using

Machine Learning," aims to predict diseases based on user-entered symptoms using machine learning algorithms. The system utilizes medical datasets containing symptom–disease relationships and applies classification algorithms such as Naive Bayes and Random Forest to generate predictions. Furthermore, the system includes an AI-powered chatbot assistant that provides users with additional health-related information and guidance.

The developed web-based platform provides an interactive interface where users can input symptoms, obtain disease predictions, and receive preventive suggestions. By combining machine learning algorithms, symptom analysis, and AI-based assistance, the proposed system aims to improve early disease awareness and support preliminary healthcare decision-making.

1.1 Machine Learning in Healthcare

Machine learning has become an important tool in modern healthcare systems for analyzing medical data and predicting diseases. By learning patterns from historical datasets, machine learning models can assist in identifying relationships between symptoms and medical conditions. Algorithms such as Naive Bayes and Random Forest are commonly used in disease prediction systems because of their high classification accuracy and efficiency in handling large datasets.

1.2 Symptom-Based Disease Prediction

Symptom-based disease prediction systems analyze the symptoms entered by users and compare them with trained medical datasets to identify possible diseases. These systems act as decision-support tools that provide preliminary diagnostic information. By utilizing machine learning techniques, symptom-driven systems can quickly process multiple symptoms and generate predictions that help users understand potential health risks.

1.3 Importance of Preventive Healthcare

Preventive healthcare plays a critical role in reducing disease risks and improving overall public health. Providing prevention recommendations along with disease prediction helps users take appropriate precautions and adopt healthier lifestyles. Integrating preventive guidance into intelligent healthcare systems ensures that users not only receive diagnostic suggestions but also learn about measures to manage and prevent diseases effectively.

2. PROPOSED SYSTEM

The proposed system aims to develop an intelligent healthcare support platform capable of predicting possible diseases based on user-provided symptoms and providing preventive recommendations. The system utilizes machine learning techniques to analyze symptom patterns and identify potential diseases from a trained medical dataset. The overall framework integrates data preprocessing, machine learning model training, symptom-based prediction, and preventive healthcare guidance.

The proposed system is designed as a web-based application that allows users to interact with the system through a simple interface. Users can enter symptoms, after which the system processes the information using trained machine learning models to generate disease predictions. Additionally, the system provides preventive recommendations and integrates an AI-powered chatbot to assist users with health-related queries. The architecture ensures efficient data processing and accurate disease prediction while maintaining a user-friendly interface.

2.1 Medical Dataset Collection

The proposed system utilizes a medical dataset containing a large number of symptoms and their corresponding diseases. The dataset serves as the primary source for training and testing the machine learning models. Each record in the dataset consists of multiple symptom attributes and the associated disease label. This structured dataset enables the machine learning algorithms to learn patterns between symptoms and diseases.

The dataset is divided into two parts: training data and testing data. The training dataset is used to build the machine learning model, while the testing dataset is used to evaluate the model's performance and prediction accuracy.

2.2 Data Preprocessing

Data preprocessing plays a crucial role in preparing the dataset for machine learning analysis. The raw medical dataset often contains redundant or inconsistent information that must be processed before training the model. The preprocessing stage includes data cleaning and data transformation processes.

During data cleaning, missing values and inconsistencies in the dataset are handled to ensure that the model receives accurate and consistent information. Data transformation converts symptom information into a structured feature format that can be used by machine learning algorithms. This step helps improve the performance and reliability of the prediction model.

2.3 Feature Extraction and Symptom Representation

In the proposed system, symptoms entered by users are converted into a feature vector representation. Each symptom corresponds to a specific feature in the dataset. When a user selects symptoms, the system converts them into a binary feature vector where the presence of a symptom is represented by 1 and the absence of a symptom is represented by 0.

This feature vector is then provided as input to the trained machine learning model. The feature representation allows the model to analyze symptom patterns efficiently and generate accurate disease predictions.

2.4 Machine Learning Model Implementation

The proposed system implements machine learning algorithms such as Naive Bayes and Random Forest Classifier for disease prediction. These algorithms are widely used for classification tasks due to their ability to handle structured datasets and provide reliable prediction results.

The Naive Bayes algorithm is used as a probabilistic classification model that calculates the probability of diseases based on the symptoms provided. Random Forest, an ensemble learning algorithm, improves prediction performance by combining multiple decision trees to generate a more robust classification result.

Both models are trained using the preprocessed medical dataset and evaluated using performance metrics such as accuracy, precision, recall, and F1-score.

2.5 Disease Prediction and Recommendation

Once the machine learning model is trained, it is used to predict diseases based on symptoms entered by users. The system processes the symptom input, converts it into a feature vector, and provides it to the trained prediction model. The model then identifies the most probable disease associated with the given symptoms.

In addition to disease prediction, the system also provides preventive recommendations to help users manage their health condition. These recommendations are retrieved from a predefined prevention knowledge base that contains preventive measures associated with each disease.

2.6 AI-Based Health Assistance

To further enhance the usability of the system, an AI-powered chatbot assistant is integrated into the platform. The chatbot allows users to ask health-related questions and receive informative responses. The chatbot uses generative AI technology to provide explanations about diseases, symptoms, and general healthcare guidance. This feature improves user interaction with the system and provides additional support beyond disease prediction.

2.7 System Architecture

Fig. 1 illustrates the overall architecture of the proposed disease prediction system. The architecture begins with the medical dataset containing disease–symptom relationships. The dataset undergoes preprocessing steps including data cleaning and transformation. After preprocessing, symptom features are extracted and converted into feature vectors.

The machine learning models are trained using the processed training dataset to build the prediction model. When new symptom data is provided by users, it undergoes preprocessing and feature vector conversion before being passed to the trained prediction model. Finally, the system outputs the predicted disease along with preventive recommendations.

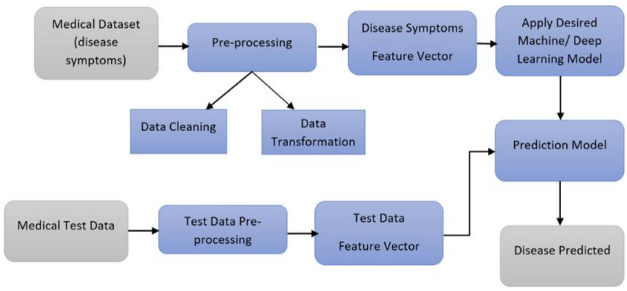

Fig. 1: System Architecture of the Proposed Disease Prediction System.

3. IMPLEMENTATION DETAILS

This section describes the practical implementation of the proposed Smart Symptom-Driven Disease Diagnosis and Prevention Recommendation System. The system is developed as a web-based application using the Django framework, integrated with machine learning models for disease prediction. The implementation includes data preprocessing, model training, symptom-based prediction, prevention recommendation, and AI chatbot integration.

3.1 Development Environment and Technologies

The proposed system is implemented using Python and the Django web framework. Python provides powerful libraries for machine learning and data processing, while Django enables the development of a dynamic web-based interface for user interaction. The system utilizes libraries such as NumPy and Pandas for data manipulation and preprocessing. Machine learning models are implemented using the Scikit-learn library, which provides efficient tools for classification and evaluation.

The user interface of the system is developed using HTML, CSS, and Bootstrap to ensure a responsive and user-friendly design. The backend logic is handled through Django views and models, which manage user authentication, prediction processing, and interaction with machine learning modules.

3.2 Dataset Preparation and Preprocessing

The medical dataset used in the system contains multiple symptoms associated with different diseases. Each row in the dataset represents a combination of symptoms and the corresponding disease label. Before training the machine learning model, the dataset undergoes preprocessing steps to ensure data consistency and quality.

During preprocessing, the disease labels are converted into numerical form using label encoding. This transformation allows machine learning algorithms to process categorical disease names efficiently. The dataset is then divided into feature variables representing symptoms and a target variable representing the disease class. The dataset is further split into training and testing subsets to evaluate model performance.

3.3 Machine Learning Model Training

The system employs two machine learning algorithms for disease classification: Naive Bayes and Random Forest Classifier. These algorithms are trained using the preprocessed training dataset to learn the relationship between symptoms and diseases.

The Naive Bayes algorithm calculates the probability of each disease based on the symptoms provided by the user. It assumes independence among features and is efficient for high-dimensional classification tasks. Random Forest, on the other hand, is an ensemble learning technique that combines multiple decision trees to improve prediction accuracy and reduce overfitting. Both models are trained using the training dataset and evaluated using the testing dataset to measure their performance.

3.4 Symptom-Based Disease Prediction

The prediction module allows users to select symptoms through the web interface. The selected symptoms are converted into a feature vector representation that matches the structure of the training dataset. Each symptom is represented as a binary value indicating whether the symptom is present or absent.

The generated feature vector is provided as input to the trained machine learning model. The model analyzes the symptom pattern and predicts the most probable disease associated with the given symptoms. The predicted disease is then displayed to the user through the web interface.

3.5 Prevention Recommendation Module

To enhance the usefulness of the system, a prevention recommendation module is integrated with the disease prediction system. After predicting the disease, the system retrieves relevant preventive measures from a predefined knowledge base. This knowledge base is implemented using a dictionary that maps diseases to their corresponding preventive recommendations.

These preventive suggestions help users understand precautionary measures and encourage early healthcare awareness. Providing preventive guidance along with prediction improves the practical value of the system.

3.6 AI Chatbot Integration

An AI-powered chatbot assistant is incorporated into the system to provide additional healthcare guidance. The chatbot uses generative AI technology to respond to user queries related to symptoms, diseases, and general medical information. Users can interact with the chatbot through the web interface by entering their health-related questions.

The chatbot analyzes the user's input and generates informative responses, helping users understand their health conditions better. This feature complements the machine learning prediction system by offering additional explanations and recommendations.

3.7 User Interface and System Integration

The entire system is integrated into a web-based platform with a user-friendly interface. The interface allows users to register, log in, access prediction modules, view algorithm performance comparisons, and interact with the AI chatbot assistant. The Django framework manages communication between the frontend interface and backend processing modules.

The integrated architecture ensures smooth interaction between data preprocessing, machine learning models, and user input. The final output presented to the user includes the predicted disease along with relevant preventive recommendations.

4. RESULTS AND PERFORMANCE ANALYSIS

The proposed Smart Symptom-Driven Disease Diagnosis and Prevention Recommendation System was developed using Python, Django, and machine learning algorithms. The system was evaluated using a medical symptom dataset containing multiple disease classes and associated symptoms. The primary objective of the evaluation was to measure the effectiveness of the machine learning models in predicting diseases accurately based on user-provided symptoms and to analyze the usability of the developed web-based interface.

The system provides an interactive web interface where users can access different functionalities such as symptom-based disease prediction, machine learning model performance comparison, and an AI-powered chatbot assistant for health-related queries. The prediction module allows users to select symptoms from a predefined list, which are then processed by the trained machine learning model to predict the most probable disease.

The developed system integrates two machine learning algorithms, namely Naive Bayes and Random Forest Classifier, to analyze and compare prediction performance. Both models were trained using the training dataset and evaluated using a testing dataset to measure accuracy, precision, recall, and F1-score. The experimental results

demonstrate that both algorithms achieve high prediction accuracy due to the structured nature of the symptom–disease dataset.

The home page of the developed system provides a clear overview of the disease prediction platform, highlighting the main features such as AI-powered prediction, symptom analysis, and accurate result generation. The interface is designed to be user-friendly and accessible, enabling users to understand the system functionality easily.

Fig. 2 illustrates the main interface of the developed disease prediction system. The page highlights the system's capability to analyze symptoms using machine learning algorithms and generate accurate disease predictions.

Fig. 2: Home Page of Disease Prediction System

The machine learning performance comparison module displays the evaluation metrics for the implemented algorithms. The comparison includes accuracy, precision, recall, and F1-score values obtained during the model evaluation process. These metrics provide insights into how well the algorithms classify diseases based on symptom data.

Fig. 3 presents the performance comparison between the Naive Bayes and Random Forest Classifier models. The results indicate that both algorithms achieved high accuracy scores, demonstrating the reliability of the proposed system in predicting diseases based on symptoms.

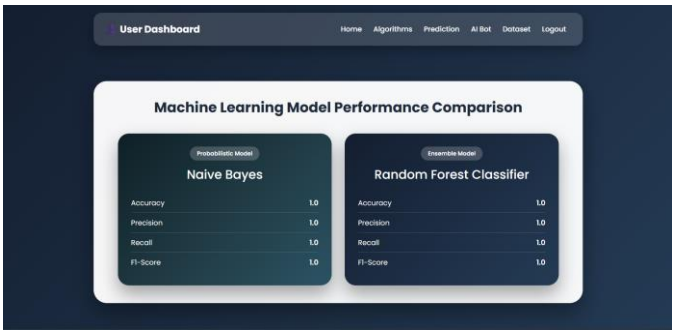

Fig. 3: Machine Learning Model Performance Comparison

In addition to disease prediction, the system integrates an AI-powered chatbot assistant that helps users obtain additional medical information and guidance. The chatbot uses generative AI technology to respond to user queries related to symptoms, diseases, and general healthcare information. This feature enhances the overall usability of the system by providing quick medical insights and improving user interaction.

Fig. 4 shows the AI chatbot interface where users can ask health-related questions and receive intelligent responses. This component complements the machine learning prediction system by offering additional explanations and recommendations.

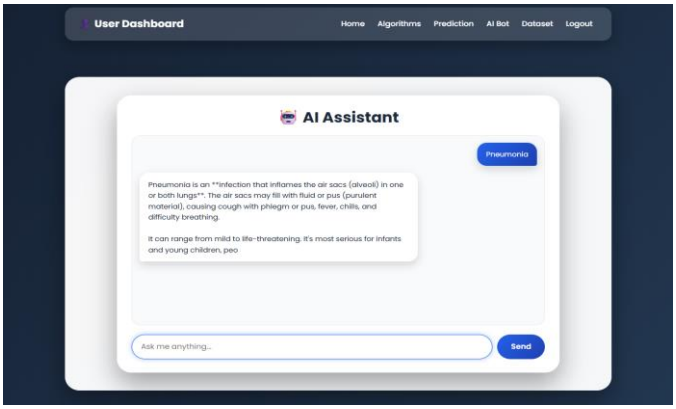

Fig. 4: AI Assistant Interface

Overall, the experimental results demonstrate that the proposed system effectively predicts diseases based on symptoms while also providing preventive recommendations and interactive medical assistance. The integration of machine learning algorithms with an AI-based chatbot improves the system's capability to support early disease identification and healthcare awareness.

5. CONCLUSIONS

In this research work, a Smart Symptom-Driven Disease Diagnosis and Prevention Recommendation System has been developed using machine learning techniques to assist users in identifying possible diseases based on their symptoms. The proposed system utilizes medical datasets containing symptom–disease relationships and applies classification algorithms such as Naive Bayes and Random Forest to predict diseases accurately. By analyzing user-provided symptoms, the system is capable of generating reliable disease predictions that can help users understand potential health conditions at an early stage.The experimental results demonstrate that the implemented machine learning models achieve high performance in terms of accuracy, precision, recall, and F1-score, indicating the effectiveness of the system in predicting diseases based on symptom data. The comparison of machine learning models further confirms that ensemble methods such as Random Forest can provide robust prediction results when trained on structured medical datasets.

In addition to disease prediction, the system also provides preventive recommendations associated with the predicted diseases. This feature enhances the practical usefulness of the system by helping users take necessary precautions and improve their health awareness. Furthermore, the integration of an AI-powered chatbot assistant allows users to obtain additional medical guidance and interact with the system in a more intuitive manner. Overall, the proposed system demonstrates the potential of combining machine learning algorithms, web-based interfaces, and AI assistance to create intelligent healthcare support tools. The system can serve as a preliminary diagnostic support platform that promotes early disease awareness and encourages users to seek professional medical consultation when necessary.

6. FUTURE WORK

Although the proposed system successfully predicts diseases based on symptoms and provides preventive recommendations, several improvements can be made in the future to enhance its functionality and accuracy. One possible enhancement is the integration of larger and more diverse medical datasets. Incorporating additional datasets containing a wider variety of symptoms and diseases can help improve the performance and reliability of the machine learning models. Another important improvement would be the use of advanced deep learning techniques for disease prediction. Algorithms such as neural networks and deep learning models can analyze more complex relationships between symptoms and diseases, potentially increasing prediction accuracy. Additionally, incorporating real-time health data from wearable devices and health monitoring systems could further improve the system's predictive capabilities.

The system can also be extended by integrating a hospital or doctor recommendation module that suggests nearby healthcare facilities based on the predicted disease. This feature would help users quickly find appropriate medical assistance when required. Furthermore, a mobile application version of the system can be developed to increase accessibility and allow users to perform symptom analysis directly from their smartphones.Future versions of the system may also include multilingual support and personalized healthcare recommendations based on user history and medical records. By integrating these enhancements, the proposed system can evolve into a more comprehensive intelligent healthcare assistant that supports early disease detection, prevention, and improved healthcare accessibility.

REFERENCES

[1] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, School of Information and ComputerSciences,2019.Available: https://archive.ics.uci.edu

[2] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," Artificial Intelligence in Medicine, vol. 23, no. 1, pp. 89–109, 2001.

[3] T. Mitchell, Machine Learning. New York, USA: McGraw-Hill Education, 1997.

[4] S. Kotsiantis, I. Zaharakis, and P. Pintelas, "Machine learning: A review of classification and combining techniques," Artificial Intelligence Review, vol. 26, no. 3, pp. 159–190, 2006.

[5] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," Machine Learning, vol. 29, pp. 103–130, 1997.

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[7] S. Ramesh, C. Yaashuwanth, and R. Rajesh, "Prediction of diseases using machine learning algorithms," International Journal of Computer Applications, vol. 181, no. 6, pp. 1–5, 2018.

[8] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," The New England Journal of Medicine, vol. 380, pp. 1347–1358, 2019.

[9] S. Deo, "Machine learning in medicine," Circulation, vol. 132, no. 20, pp. 1920–1930, 2015.

[10] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, USA: Morgan Kaufmann, 2012.

[11] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[12] W. W. Cohen, "Fast effective rule induction," in Proceedings of the 12th International Conference on Machine Learning, 1995, pp. 115–123.

[13] J. Brownlee, Machine Learning Mastery with Python. Melbourne, Australia: Machine Learning Mastery, 2016.

[14] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 3rd ed. Upper Saddle River, NJ, USA: Pearson Education, 2010.

[15] K. H. Zou, A. J. O'Malley, and L. Mauri, "Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models," Circulation, vol. 115, pp. 654–657, 2007.