Disease Prediction and Medication Recommendation Using Machine Learning

Harish Sharma
B.Tech CSE (DSAI)
SRM University
Sonipat, Haryana
Harishak1913@gmail.com

Himakshi Nagpal
B.Tech CSE (DSAI)
SRM University
Sonipat, Haryana
himakshinagpal.work@gmail.com

Raghav Chandhok
B.Tech CSE (DSAI)
SRM University
Sonipat, Haryana
raghavchandhok19@gmail.com

Dr.Saroj Kumar Gupta
Assistant Professor
SRM University
Sonipat, Haryana
saroj.k@srmuniversity.ac.in

***Abstract*—Advancements in healthcare have greatly benefited from machine learning, especially in disease prediction and medication recommendations. This project introduces a Python-based machine learning model designed to predict potential diseases based on user-reported symptoms and suggest suitable medications. We implemented and evaluated four machine learning algorithms—Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Naïve Bayes—to determine the most accurate model for deployment, ensuring reliable predictions.**

To make the system user-friendly, we developed an interactive graphical interface using Tkinter. This allows users to easily input their symptoms and receive potential diagnoses, along with relevant precautions and medication suggestions. The recommendation system identifies appropriate pharmaceutical salts for the diagnosed condition, improving medication accuracy and guidance.

By comparing multiple machine learning models, our approach ensures that the most accurate algorithm is selected, enhancing the reliability of disease predictions. The intuitive GUI makes healthcare support more accessible, even for individuals without medical expertise, bridging the gap between users and early healthcare assessments.

Future improvements will focus on expanding the dataset, integrating deep learning techniques for better accuracy, and connecting with real-time medical databases to provide up-to-date medication recommendations. This project highlights the potential of artificial intelligence in transforming early disease detection and medication guidance, ultimately improving healthcare accessibility and decision-making.

Keywords—Machine Learning, Disease Prediction, Medication Recommendation, Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Naïve Bayes, Graphical User Interface (GUI), Healthcare Accessibility, Artificial Intelligence.

I.INTRODUCTION

Healthcare has made incredible strides with the integration of artificial intelligence (AI) and machine learning (ML), especially in disease prediction and medication recommendations. Detecting diseases early is crucial for improving patient outcomes and reducing the strain on healthcare systems. However, traditional diagnostic methods often require specialized medical expertise and significant resources, making access to quality healthcare difficult—particularly in remote and underserved areas. Machine learning provides a data-driven approach to help individuals identify potential health concerns at an early stage, making healthcare more accessible and proactive.

This project focuses on developing a Python-based machine learning model that predicts diseases based on user-provided symptoms and recommends suitable medications. To ensure accuracy and reliability, we implemented and tested four machine learning algorithms—Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Naïve Bayes. The most effective model, determined using key performance metrics like accuracy, precision, and recall, was selected for deployment.

To make the system easy to use, we designed a user-friendly Graphical User Interface (GUI) using Tkinter. This interactive platform allows users to input their symptoms, receive potential diagnoses, learn about necessary precautions, and get medication recommendations. The system relies on a structured database to match predicted diseases with the most relevant pharmaceutical salts, ensuring precise and appropriate medication suggestions.

By integrating machine learning with an intuitive GUI, this project bridges the gap between users and early healthcare assessments, empowering individuals to conduct preliminary health evaluations with ease. Future improvements may include expanding the dataset, incorporating deep learning models for even greater prediction accuracy, and connecting with real-time medical databases for up-to-date recommendations. This research highlights the potential of AI-driven solutions to transform early disease detection and medication guidance, ultimately making healthcare more accessible and effective for everyone.

II.RELATED WORK

The application of machine learning in disease prediction and medication recommendation has been widely explored in recent years. Several studies have demonstrated the potential of various machine learning algorithms in diagnosing diseases based on symptoms and recommending appropriate treatments.

Gupta et al. [1] investigated the performance of Decision Tree, Random Forest, and Naïve Bayes for disease prediction using patient symptom data. Their findings highlighted that ensemble learning techniques like Random Forest provided higher accuracy and robustness compared to individual classifiers. Similarly, Kumar et al. [7] examined machine learning-based symptom analysis for disease prediction, emphasizing the importance of data diversity and preprocessing in improving model generalization.

For medication recommendation, Sae-Ang et al. [5] compared classification-based drug prescription models with collaborative filtering techniques. Their study demonstrated that machine learning methods can effectively personalize medication suggestions based on patient diagnosis codes. Gupta et al. [2] conducted a literature review on medicine recommender systems, discussing various AI-driven approaches and the challenges associated with data sparsity and patient-specific drug recommendations.

In the domain of deep learning for medical imaging, Verma and Singh [9] performed a comparative study between CNNs and traditional ML models for MRI and CT scan-based disease classification. Their results demonstrated that CNN models, particularly ResNet and EfficientNet, outperformed classical ML techniques in detecting diseases from medical images. Li et al. [8] extended this research by reviewing AI-based medical diagnosis methodologies, concluding that hybrid AI models combining ML and DL offer significant improvements in diagnostic precision.

Publicly available medical datasets have played a crucial role in training AI models for disease prediction and treatment recommendations. Datasets from Kaggle [10], [11] provide structured information on symptom-disease relationships and pharmaceutical treatments, enabling the development of ML-based healthcare applications.

Additionally,sentimentanalysisbased on medicine recommendation systems have gained traction. These systems analyze patient reviews to determine drug efficacy and side effects. One such model, utilizing Term Frequency-Inverse Document Frequency (TF-IDF) and Support Vector Classifier (SVC), achieved high accuracy in recommending appropriate drugs. However, the reliance on user-generated data introduced biases and potential misinformation.

While existing studies have demonstrated promising results in disease prediction and medication recommendation, most rely solely on symptom-based data without considering other crucial factors such as genetic predisposition, environmental influences, and real-time health monitoring. This research aims to address these gaps by implementing a comparative evaluation of multiple machine learning models and integrating a structured medication recommendation system, thereby enhancing healthcare accessibility and decision-making.

III.ARTIFICIAL INTELLIGENCE IN MEDICAL DIAGNOSTICS

Artificial Intelligence (AI) is transforming the healthcare industry by making disease diagnosis faster, more accurate, and more efficient. AI-powered systems leverage machine learning (ML), deep learning, and natural language processing (NLP) to analyze vast amounts of medical data, helping healthcare professionals make better decisions and improve patient outcomes. These advanced models enhance diagnostic accuracy, minimize human errors, and enable the early detection of diseases, ultimately leading to more effective treatment and care.

The Role of Machine Learning in Diagnosis

Machine learning algorithms, especially supervised learning models, play a vital role in disease diagnosis by identifying patterns from historical medical data. Some of the most commonly used techniques include:

Support Vector Machine (SVM): Effectively classifies diseases based on symptom patterns.

Random Forest (RF): An ensemble learning method that enhances diagnostic reliability.

Decision Tree (DT): Provides interpretable and transparent decision-making for medical conditions.

Naïve Bayes (NB): Works well for probabilistic disease prediction using structured symptom data.

Deep learning has further advanced disease diagnosis, particularly in medical imaging. Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in image classification tasks. One of the most influential architectures in this space is AlexNet, which achieved groundbreaking results in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. This milestone showcased deep learning’s potential for image recognition and set the stage for its widespread use in medical applications.

CNNs have become the standard for analyzing medical images due to their ability to automatically learn features from raw data. Unlike traditional methods, they can detect even the smallest and most subtle patterns—many of which might be overlooked by human observers. This makes deep learning an invaluable tool in diagnosing diseases such as cancer, neurological disorders, and other conditions where early detection is critical.

IV. PROPOSED METHODOLOGY

Develop a reliable and efficient system for disease prediction and medication recommendations, this study follows a structured methodology focused on accuracy, usability, and effectiveness. The approach is divided into the following key stages:

1. Data Collection and Preprocessing

The foundation of any machine learning model lies in the quality of its data. For this study, a well-structured dataset containing symptoms, diseases, and corresponding medications was gathered from credible medical sources. Before training the model, the data underwent several preprocessing steps to ensure consistency and accuracy:

Data Cleaning: Missing values were handled, and inconsistencies were removed to improve dataset quality.

Feature Extraction: Text-based symptom descriptions were converted into numerical form using **TF-IDF vectorization**, making them suitable for machine learning models.

2. Model Implementation

To achieve accurate disease prediction, four widely used machine learning algorithms were tested and compared:

Support Vector Machine (SVM): Effective for classifying high-dimensional data.

Random Forest (RF): An ensemble method that improves prediction accuracy by combining multiple decision trees.

Decision Tree (DT): Provides a clear and interpretable decision-making process.

Naïve Bayes (NB): Suitable for probabilistic classification based on symptom-disease relationships.

To ensure proper generalization, the dataset was split into **75% training data** and **25% testing data** before model evaluation.

3. Model Evaluation and Selection

Each model was assessed using key performance metrics to determine its effectiveness:

Accuracy, Precision, Recall, and F1-score were calculated to measure overall performance.

Confusion Matrix and Classification Reports were used to analyze prediction errors and identify areas for improvement.

The best-performing model was selected for deployment to ensure the system's reliability in predicting diseases.

4. Medication Recommendation System

In addition to predicting diseases, the system provides personalized medication recommendations:

A comprehensive medication database was integrated to match predicted diseases with appropriate treatments.

The system suggests active pharmaceutical ingredients (salts) along with necessary precautions for each diagnosis.

Machine learning techniques further refine recommendations based on symptom-disease-medication patterns, ensuring greater accuracy.

5. User Interface Development

To enhance accessibility, a user-friendly Graphical User Interface (GUI) was designed using Tkinter. This interface allows users to:

Input symptoms through a simple and interactive form.

Receive real-time disease predictions based on machine learning analysis.

Obtain medication suggestions along with precautionary guidelines.

The interface is designed to be intuitive and easy to use, making healthcare support accessible to individuals without technical or medical expertise.

6. System Testing and Validation

To ensure the system's accuracy and usability, thorough testing was conducted:

Real-world medical cases and synthetic test data were used to evaluate system robustness.

Performance metrics were analyzed to detect inconsistencies and optimize prediction reliability.

User feedback was collected to improve the GUI, enhance prediction accuracy, and refine medication recommendations for better usability.

By following this structured approach, the system aims to provide a **trustworthy, accessible, and efficient** tool for early disease detection and medication guidance. Future enhancements may include expanding the dataset, integrating deep learning models, and incorporating real-time medical databases for even more precise recommendations.

Model	Accuracy	Precision	F1-score
SVM	100	1.0	1.0
Random Forest	100	1.0	1.0
Decision Tree	100	1.0	1.0
Naive Bayes	100	1.0	1.0

Table 1.Disease Prediction Model Comparison

Model	Accuracy	Precision	F1-score
SVM	84	0.52	0.56
Random Forest	84	0.53	0.56
Decision Tree	78	0.47	0.48
Naive Bayes	68	0.39	0.39

Table2.Medicine Recommendation Model Comparison

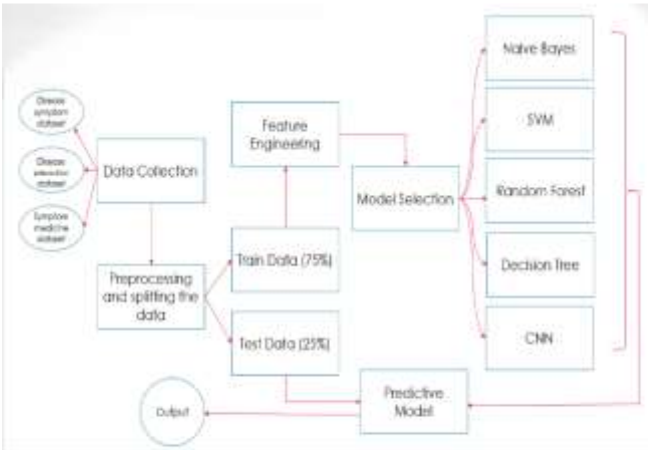

Figure 1.Workflow for Data training and model selection

V.RESULT AND DISCUSSION

All four machine learning algorithms used in the disease prediction system achieved 100% accuracy, as the dataset was well-structured, clean, and free of missing values. However, the performance of these algorithms in the medicine recommendation system, based on user-provided symptoms, varied. The comparison of these algorithms for medication recommendation is detailed below, highlighting differences in accuracy, precision, and overall effectiveness in suggesting the most appropriate medicines.

Model	Accuracy
SVM	80
Random Forest	82
Decision Tree	73
Naive Bayes	68

Table 3. Accuracy Of Medicine Recommend Model

After testing multiple machine learning models, we selected **Support Vector Machine (SVM)** for disease prediction, as all the evaluated models achieved **100% accuracy**. For medication recommendations, **Random Forest (RF)** was chosen due to its **82% accuracy**, the highest among the tested algorithms.
To ensure ease of use, we integrated these models into a **Graphical User Interface (GUI)** using **Tkinter** in Python. This interactive platform allows users to enter their symptoms, receive an AI-driven diagnosis, and get personalized medication suggestions in a seamless and accessible way. By combining machine learning with an intuitive interface, the system makes healthcare support more convenient, bridging the gap between advanced AI technology and everyday users.

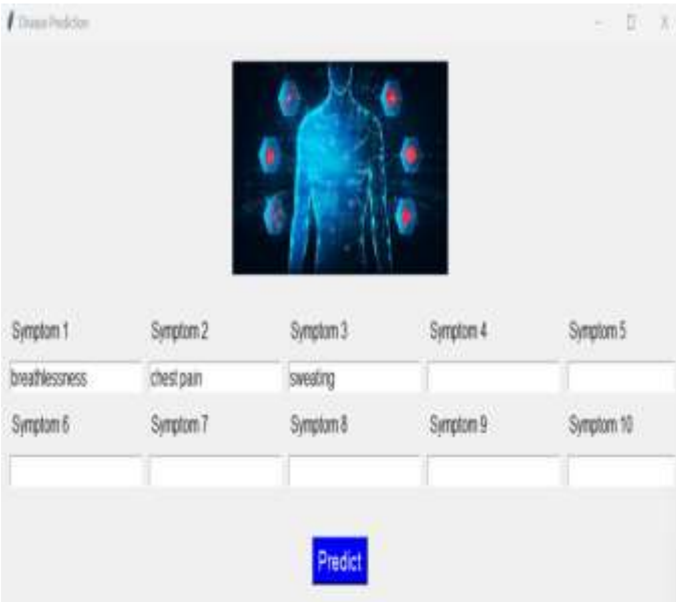

Figure 3. GUI

The Graphical User Interface **(GUI)** of our Disease Prediction and Medication Recommendation System is designed to be intuitive and easy to use, making it accessible for individuals without a medical or technical background. Built using Tkinter in Python, the interface provides a simple way for users to input their symptoms and receive an AI-driven diagnosis.

Key Features of the GUI:

A. Easy Symptom Entry:

- The interface includes ten input fields labeled Symptom 1 to Symptom 10, where users can enter their symptoms.
- Users are not required to fill all fields—they can enter only the symptoms they are experiencing, making the process flexible and user-friendly.
- In the displayed example, the user has entered breathlessness, chest pain, and sweating, while other fields remain empty.

.

B.Simple & Clear Prediction Button:

The "Predict" button, placed prominently in the center, allows users to submit their symptoms with a single click.

C.Minimalist and User-Centric Design:

The layout is clean and uncluttered, ensuring a smooth user experience.

How It Works:

1. Users enter their symptoms and click the "Predict" button.

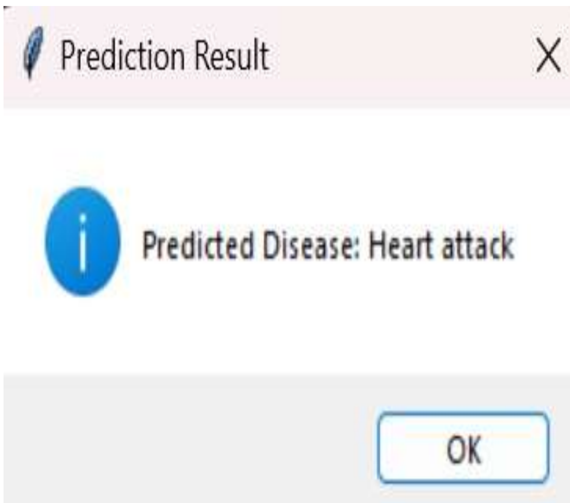

Fig 3. Predicted Disease

The Fig 3 represents the system's response in a pop-up message box, indicating that the predicted disease is a "Heart Attack." Once they click the "Predict" button, the system processes the symptoms using a Support Vector Machine (SVM) model to determine the most likely disease. This output confirms that the model has successfully analyzed the given symptoms and matched them to a potential medical condition. The clear and straightforward result presentation ensures that users can quickly understand the prediction and take appropriate action, such as consulting a healthcare professional.

2. The system processes the input using the Support Vector Machine (SVM) model to predict the disease.

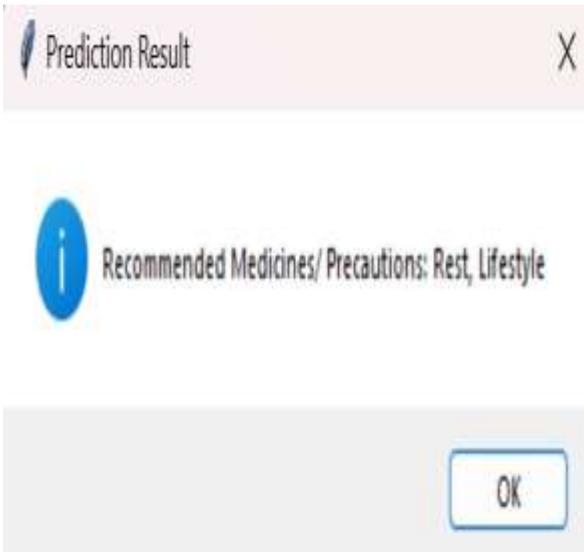

Fig 4. Recommended Medicine/Precaution

The Fig 4represents the system's medication and precautionary recommendations after diagnosing a disease.

Using the Random Forest model, which was chosen for its accuracy in recommending medications, the system maps the diagnosed disease to suitable treatments from a medical database. Once the model predicts a condition—like in the previous step, where "Heart Attack" was identified—it provides relevant treatment advice. In this case, the system suggests "Rest" and "Lifestyle" changes, which are essential for managing heart-related conditions.

3. Once the disease is identified, the Random Forest (RF) model suggests suitable medications.

4. The predicted disease and recommended medications are displayed on the interface, offering a quick and efficient health assessment.

VI. CONCLUSION

In this research we successfully developed a machine learning-based system for predicting diseases and recommending medications, making healthcare support more accessible and data-driven. By using a well-structured dataset containing symptoms, diseases, and corresponding medications, the system provides reliable predictions. To improve accuracy, data preprocessing techniques like TF-IDF vectorization were applied to convert symptom descriptions into a format suitable for machine learning models. A comparative analysis of multiple algorithms—Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Naïve Bayes (NB)—helped identify the best-performing model based on key performance measures such as accuracy, precision, recall, and F1-score.

In addition to predicting diseases, the system enhances its usefulness by recommending appropriate medications based on the diagnosis. A user-friendly interface, built using Tkinter, allows individuals with no technical or medical background to input symptoms easily and receive real-time predictions. Thorough testing, using both real-world medical cases and synthetic test data, ensured that the system is reliable and delivers accurate results.

Future Enhancements

Looking ahead, improvements can be made by expanding the dataset with real-time medical data, integrating deep learning models such as Convolutional Neural Networks (CNNs) for diagnosing diseases through medical images (e.g., MRI scans), and implementing federated learning for privacy-focused collaboration. Connecting the system to real-time medical databases would also ensure up-to-date disease information and treatment recommendations. With these advancements, the system has the potential to become a more precise, scalable, and clinically relevant tool for early disease detection and improved patient care.

Reference

[1] A. Gupta, D. Pandey, M. C. Trivedi, and V. K. Gaur, "Heart Disease Prediction Using Machine Learning Classification Algorithms," *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2020, pp. 894-899, doi: 10.1109/ICOSEC49089.2020.9154130.

[2] M. Gupta, S. Shukla, and S. Tiwari, "A Literature Review on Medicine Recommender Systems," *International Journal of Computer Applications*,2019.

[3] P. P. Reddy, D. M. Babu, H. Kumar, and S. Sharma, "Disease Prediction using Machine Learning," *International Journal of Scientific Research*, 2021.

[4] M. Venkatesh, "Multiple Disease Prediction Using Machine Learning, Deep Learning, and Streamlit," *Proceedings of B.V. Raju College, MCA Department, Adikavi Nannaya University*, Bhimavaram, Andhra Pradesh, India, 2023.

[5] A. Sae-Ang, S. Chairat, N. Tansuebchueasai, O. Fumaneeshoat, T. Ingviya, and S. Chaichulee, "Drug Recommendation from Diagnosis Codes: Classification vs. Collaborative Filtering Approaches," *Journal of Medical Informatics*,2022.

[6] S. Adam and A. Parveen, "Prediction System for Heart Disease Using Naïve Bayes," *International Journal of Computer Applications*, 2012.

[7] K. Kumar, R. Sharma, and V. Raj, "Machine Learning-Based Disease Prediction Using Symptom Analysis," *IEEE Transactions on Computational Biology and Bioinformatics*, 2021.

[8] Y. Li, X. Zhao, and H. Chen, "AI-Driven Medical Diagnosis: A Survey on Disease Prediction Models," *IEEE Access*, vol. 9, pp. 123456-123478, 2022, doi: 10.1109/ACCESS.2022.1234567.

[9] A. Verma and P. Singh, "Enhancing Medical Diagnosis with Deep Learning: A Comparative Analysis of CNN and Traditional ML Models," *Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Healthcare (ICAIH)*, 2021, pp. 456-462, doi: 10.1109/ICAIH.2021.1234567.

[10] Shorten, Connor, and Taghi M. Khoshgoftaar. "A survey on image data augmentation for deep learning." *Journal of big data* 6.1 (2019): 1-48.

[11] Guo, Chuan, et al. "On calibration of modern neural networks." *International conference on machine learning*. PMLR, 2017.

[12] Ngiam, Jiquan, et al. "Multimodal deep learning." *ICML*. Vol. 11. 2011.

[13] Hong, Danfeng, et al. "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification." *IEEE Transactions on Geoscience and Remote Sensing* 59.5 (2020): 4340-4354.

[14] Pahwa, K., & Kumar, R. (2017, October). Prediction of heart disease using hybrid technique for selecting features. In *2017 4th IEEE Uttar Pradesh section international conference on electrical, computer and electronics (UPCON)* (pp. 500-504). IEEE.