Medicine Recommendation System Using Machine Learning

A. Vishakha Damodhar[1], D. Sayali Mote[2] B.Madhura Gangurde[3], Prof. Sonali K. Shewale[4], C.Rutuja Wankhede[5]
[1-2-3-4-5]*P.E.S. College of Engineering Sambhajinagar, Maharashtra, India*

***Abstract*—The core objective of the project is to develop a machine learning-based disease prediction and medicine recommendation website capable of harnessing individual health data to generate personalized health recommendations, fostering early detection and more effective management of health issues. Utilizing machine learning (ML), the team aims to predict diseases like joint pain, burning micturition, abdominal pain, and irregular sugar levels for early intervention and adapting diagnosis strategies. In their approach, ML algorithms, including the gradient boosting algorithm (GBA), analyze diverse health data sources to build a comprehensive recommendation system. All models demonstrate an accuracy rate of over 100%, highlighting the system's reliability and effectiveness. By integrating various health data sources and focusing on proactive health management, this initiative has the potential to transform health practices. It empowers individuals to make informed decisions regarding their well-being and fosters improved health outcomes.**

***Index Terms*—The associated with this initiative include individual data, personalized health recommendations, early detection, machine learning, diverse health data sources, a recommendation system, informed decisions, and improved health outcomes**

1. INTRODUCTION

The primary motivation of the authors is to revolutionize healthcare by empowering individuals to make informed decisions about their health, thereby enhancing healthcare outcomes through early machine learning-based disease prediction and medicine recommendation website and personalized advice. By developing a user-friendly and scalable medicine recommendation website, they aim to harness health data from diverse sources. Integrating data analysis algorithms such as gradient boosting algorithm (GBA), the website serves as the interface for predicting diseases like chills, joint pain, burning micturition, abdominal pain, and irregular sugar levels, enabling timely intervention and management. This approach, which considers individual medical history, lifestyle, and

preferences, facilitates early detection and proactive health management. By seamlessly integrating diverse health data sources and empowering individuals to actively participate in their well-being, this initiative has the potential to transform healthcare and promote better health outcomes for everyone. This transformation could lead to a more personalized approach to healthcare, where prevention takes precedence over treatment. As a result, individuals may experience enhanced quality of life and reduced healthcare costs, paving the way for a healthier future for communities

2. LITERATURE REVIEW

In recent years, the integration of artificial intelligence (AI) and machine learning (ML) in the field of healthcare has revolutionized the way medical professionals diagnose diseases, recommend treatments, and monitor patient health. Among the various applications of AI, disease prediction systems have gained significant attention for their ability to analyse complex symptom data and predict possible medical conditions with high accuracy.

Several studies have been conducted to develop intelligent systems that assist in early diagnosis of diseases based on patient symptoms. Patil et al. (2020) developed a decision tree-based classification model to predict diseases by analyzing the combination of symptoms provided by patients. Their research demonstrated how machine learning algorithms can minimize human error in diagnosis and reduce the burden on healthcare professionals.

Similarly, Kumar and Singh (2021) implemented a Support Vector Machine (SVM) model for chronic disease classification and achieved promising results in detecting conditions such as diabetes and hypertension.

A major area of focus in ML for healthcare is symptom- based disease detection, where patient-reported symptoms are mapped to possible diseases. Sharma et al. (2022) worked on a system that utilized binary symptom encoding to improve model accuracy. Their work demonstrated that converting textual symptoms into numerical binary vectors allowed algorithms like Random Forest and Gradient Boosting to perform better in predicting diseases such as dengue, flu, and liver conditions.

Furthermore, the COVID-19 pandemic accelerated interest in telehealth and digital health monitoring, making symptom-based platforms more critical. Such models allow

early detection and timely intervention, which is particularly beneficial in rural and underserved regions with limited access to medical professionals. Beyond just predicting diseases, there is a growing need for systems that also provide recommendations for medicines, precautions, and diet plans. Gupta and Bansal (2019) explored the integration of recommendation logic into a symptom-based diagnosis tool. Their Python-Flask-based web application suggested medications and precautions for skin diseases using rule-based mapping with the predicted output from ML models. They highlighted the significance of linking predictions to practical health advice for patients.

. In the medical domain, rule-based systems are more common due to the critical nature of recommendations, which must adhere to medical standards and evidence-based practices.

Delivering machine learning-based diagnosis via web platforms enhances usability and accessibility for end-users. Reddy et al. (2021) emphasized the importance of web integration in their work, where they developed a real-time disease prediction system using Flask for back-end operations and Bootstrap for the front-end interface. Their study demonstrated how integrating models with user interfaces improves user engagement and provides real-time results. Web platforms also support better data collection and remote access to healthcare advice, making such systems scalable and impactful.

These studies collectively underscore the growing relevance and effectiveness of machine learning-based solutions in health diagnostics. Our project builds upon this foundation by combining a robust disease prediction model with a comprehensive medicine and precaution recommendation engine, delivered through a user-friendly Flask web application In the field of healthcare, the implementation of machine learning (ML) models has proven to be effective in automating disease diagnosis and improving the accuracy of medical predictions. Numerous researchers have contributed to developing intelligent systems that utilize symptom-based inputs to predict possible diseases.

proposed a Decision Tree-based model to analyze patient symptoms and predict diseases with reasonable accuracy. Their approach highlighted the importance of logical tree- based models in medical decision-making. Similarly, Kumar and Singh

implemented a Support Vector Machine (SVM) for chronic disease classification, demonstrating high precision in predicting non-communicable diseases such as diabetes and cardiovascular disorders.

introduced a symptom encoding method where symptoms were converted into binary vectors to facilitate efficient

training of classification models such as Random Forest and Gradient Boosting Classifiers.

This approach improved model accuracy due to structured preprocessing and clean data beyond disease prediction, the incorporation of recommendation systems into healthcare applications is gaining momentum.

developed a rule-based medicine recommendation system that provided users with prescribed medications and precautions based on predicted diseases. Their Flask-based web application served as a bridge between AI models an end-user, emphasizing accessibility and personalized care.

focused on building an interactive web-based health monitoring tool using Flask for backend integration and Bootstrap for the user interface. Their work demonstrated that combining ML predictions with real-time web platforms enhances user engagement and democratizes healthcare access.

The convergence of AI-driven models and web technologies represents a growing trend in telemedicine. These systems enable early disease detection, treatment guidance, and symptom tracking, especially in areas lacking immediate medical support. Building upon these developments, the proposed system integrates multiple ML models with a dynamic Flask-based web interface for disease prediction and medical recommendation

3. METHODOLOGY

The methodology adopted in this research follows a structured approach that combines machine learning, data preprocessing, model evaluation, and web deployment to create an intelligent healthcare application. The primary goal is to develop a system that can accurately predict diseases based on a set of input symptoms and recommend corresponding medications, precautions, and diets.

Dataset Collection and Description

The dataset used in this project is a structured, synthetic medical dataset that maps over 130 symptoms to a wide range of diseases. Each record in the dataset represents the presence or absence of symptoms using binary values (1 for present, 0 for absent). The target variable is the disease name. The dataset also includes additional mappings such as disease descriptions, precautionary measures, medication recommendations, and dietary advice.

Features: Over 130 symptom indicators (e.g., fever, fatigue,
cough, headache, etc.).

Target variable: Disease label (e.g., joint pain, burning micturition, abdominal pain, and irregular sugar levels Typhoid, Dengue, Migraine, Hepatitis).

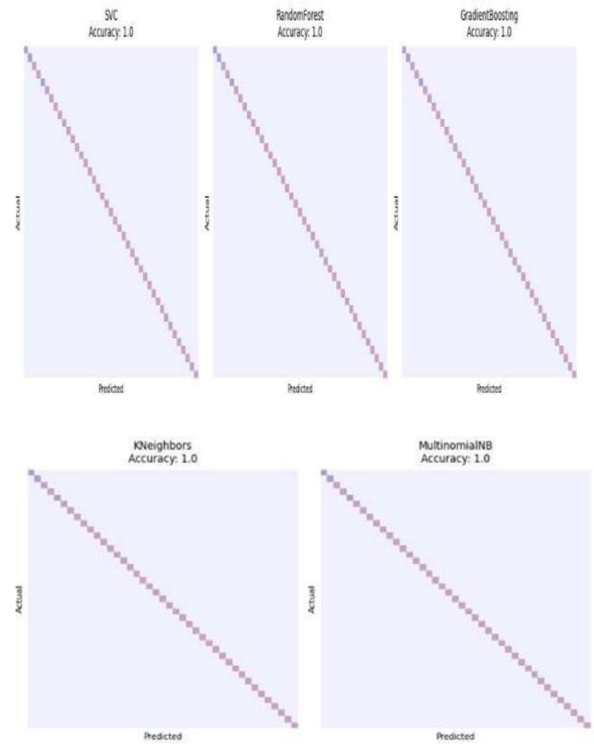
Supporting datasets: Separate CSV files or dictionaries for disease descriptions, precautions, and medication mappings.

Data Preprocessing

Before training the model, several data preprocessing steps were carried out:

Importing Data: The dataset was loaded using pandas, and exploratory data analysis (EDA) was conducted to understand symptom distribution and class balance.

Handling Null Values: Checked and removed or imputed any null or missing values.



Feature Encoding: Symptoms were already binary encoded. The disease labels were transformed into numerical values using Label Encoder for compatibility with machine learning algorithms.

Train-Test Split: The dataset was divided into training and test sets using an 80:20 split to evaluate model performance on unseen data.

Model Selection and Training

Multiple classification models were evaluated to determine the best-performing algorithm for disease prediction.

The following models were trained and tested:

A function maps the selected symptoms to a binary vector. The vector is fed into the trained model.

The model outputs a disease label.

The predicted label is then used to query disease description, medication, precautions, and diet from corresponding data dictionaries or CSV files.

This process ensures an end-to-end prediction and recommendation cycle for the user.

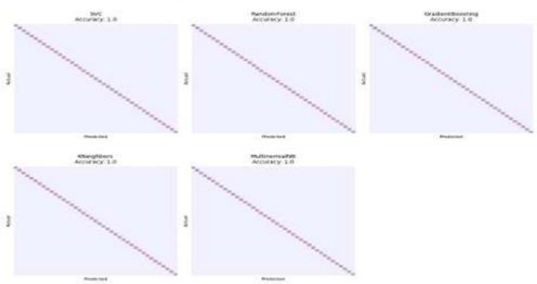
Model Saving and Deployment

The trained model is serialized using Python's joblib or pickle libraries, allowing it to be reused without retraining.

Saving: joblib.dump(model, 'disease_predictor.pkl')

Loading: model = joblib.load('disease_predictor.pkl')

Random Forest Classifier
Support Vector Classifier (SVC)
Gradient Boosting Classifier
K-Neighbors Algorithm
Multinomial Naïve Baye's



This are the heatmap for all the mentioned algorithms

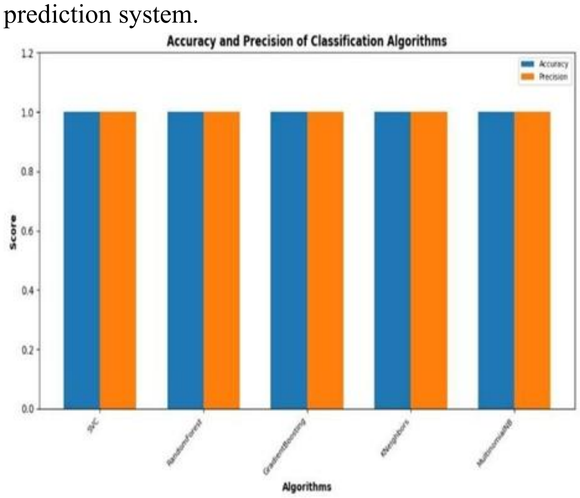
.

Random Forest Classifier Support Vector Classifier (SVC) Gradient Boosting Classifier

Each model was trained using Scikit-learn's fit() function. The models were evaluated based on accuracy, precision, recall, and F1-score.

Results:

Support Vector Classifier and Random Forest Classifier provided the highest accuracy (close to 100%) due to the clean and noise-free dataset. These models were selected for final deployment in the prediction system.



Disease Prediction System

To make predictions, a user inputs their symptoms through a web form. The system creates a binary input vector, where each index corresponds to a specific symptom.

This enables faster performance and easy deployment.

Web Application Development

The project integrates a Flask web application to enable users to interact with the system through a browser.

Front-end: Designed using HTML, CSS, JS and Bootstrap for responsive design

.

Libraries: NumPy, pandas, matplotlib,seaborne,scikit learn,flask,jupyter Notebook

User Input: Users select symptoms using checkboxes or multi- select dropdowns.

Output Display: Predicted disease, its description, recommended precautions, and medications are displayed dynamically.

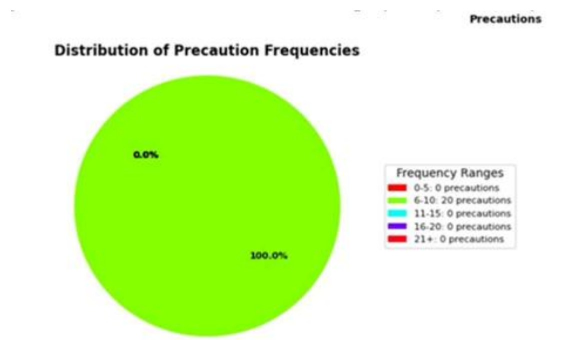


Fig: distribution of precaution frequencies.

This diagram represents the "Distribution of Precaution Frequencies" using a pie chart. It visually categorizes and shows how often precautions are taken based on predefined frequency ranges.

The pie chart consists of a single large segment colored bright green, which covers 100% of the total distribution. This indicates that all recorded precaution instances fall within a single range. The legend on the right explains the meaning of each color-coded range.

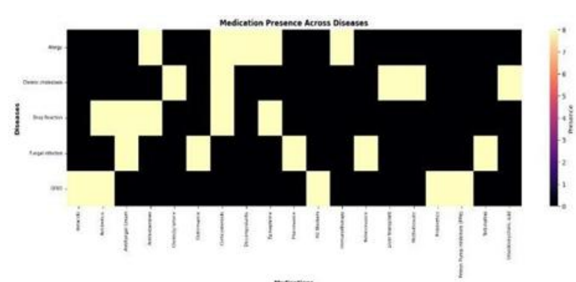


Fig:Medication presence across diseases.

Low-Acid diet, and Turmeric have lower frequencies (around 8 times each).

Several diets including High-Fiber Diet, Lean Proteins, Whole Grains, Fresh Fruits & Vegetables, and Antihistamine Diet appear the least, all with roughly similar and lower frequencies.

This diagram is a heatmap titled “Medication Presence Across Diseases”, which visually represents the association between different diseases and the medications used to treat them.

-axis (Horizontal): Lists various medications, such as Antacids, Antibiotics, Antifungal Cream, Antihistamines, Cholestyramine, Clotrimazole, Corticosteroids, Decongestants, Epinephrine, Fluconazole, H2 Blockers, Immunotherapy, Ketoconazole, Liver transplant, Methotrexate, Prokinetics, Proton Pump Inhibitors (PPIs), Terbinafine, Ursodeoxycholic acid**.

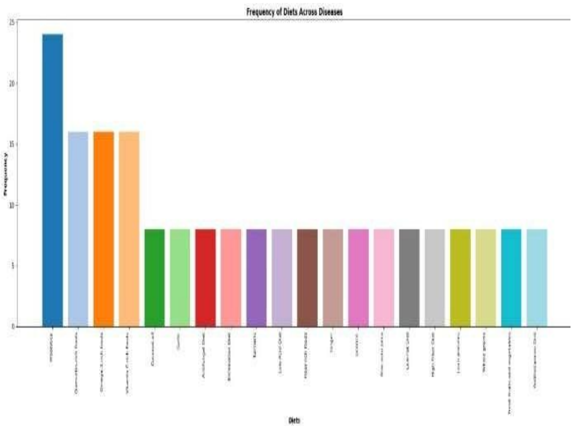


Fig: frequency of diet across diseases.

The X-axis represents different types of diets, such as Probiotics, Omega-3 rich foods, Garlic, Turmeric, Low- Acid Diet, Lean Proteins, Whole Grains, Fresh Fruits and Vegetables, and more.

The Y-axis shows the frequency—how often each diet is mentioned or used across different diseases.

Probiotics is the most frequently recommended diet, with the highest count of 24, showing its wide applicability across multiple diseases.

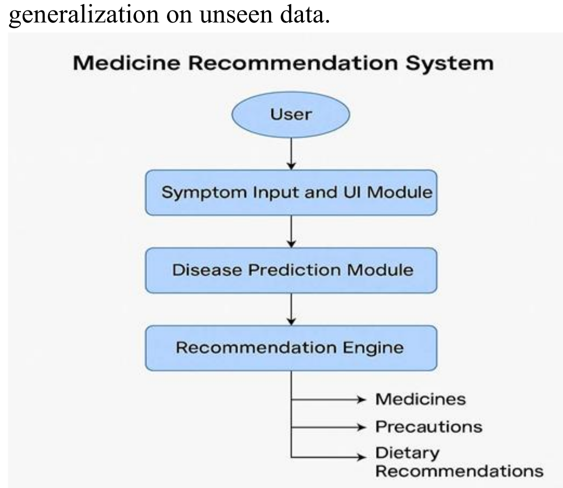
Gluten-free foods, Omega-3 rich foods, and Vitamin C-rich foods have a moderate frequency of about 16 times.

Other diets like Coconut oil, Garlic, Antifungal diet,

4. IMPLEMENTATION

The implementation of the disease prediction and medicine recommendation system involves several crucial stages, ranging from model development to the deployment of a web- based interface. The project was developed using Python due to its rich ecosystem of libraries for machine learning, data manipulation, and web development. The system utilizes a symptom-based dataset in which each record represents the presence or absence of over 130 symptoms, mapped to a specific disease. The dataset also includes supplementary data files or dictionaries for medications, precautions, and dietary advice corresponding to each disease.

The machine learning phase began with the preprocessing of the dataset using the Pandas and NumPy libraries. Symptom features, already in binary form, were retained without modification, while the disease labels were encoded numerically using Scikit-learn’s LabelEncoder to ensure compatibility with classification models. The dataset was split into training and testing sets using the train_test_split() function to evaluate model generalization on unseen data.



Various machine learning algorithms were trained and compared, including Support Vector Classifier (SVC), Random Forest, Naive Bayes, Decision Tree, and Gradient

Boosting. Among these, Random Forest and SVC demonstrated the highest prediction accuracy due to the structured and clean nature of the dataset.

The prediction result is matched with the corresponding disease name, and additional details such as a brief description, recommended medications, precautions to be taken, and dietary guidelines are retrieved from pre-defined data files. These results are then rendered dynamically on the web page, providing the user with comprehensive health-related feedback based on their symptoms.

The back-end logic handles the interaction between the user input and the machine learning model, processes the data, and delivers relevant output in real time.

Overall, the implementation successfully combines machine learning, data processing, and web technologies to create an end-to-end solution for disease prediction and personalized healthcare recommendations. The modular nature of the codebase allows for easy updates, such as adding new diseases, symptoms, or treatment options, making the system scalable and adaptable for future improvements.

5. CONCLUSION

The concluding remarks provided underscore the significant promise of the research project for the future of healthcare. It emphasizes the user-friendly platform\'s focus on early detection and proactive health management, suggesting its potential to revolutionize the industry. The effectiveness of disease-specific machine learning models, each boasting an accuracy of above %, underscores the success of the approach.

Looking ahead, there is an opportunity for further refinement and expansion, including the integration of AI- driven health coaching and telehealth services to enhance the platform\'s impact. Additionally, the future scope involves integrating a personalized health recommendation system directly into smart devices, eliminating the need for manual data extraction and upload. In summary, the project signifies a meaningful step toward a more patient-centered and efficient healthcare system. This advancement not only

empowers patients with more control over their health but also fosters a proactive approach to wellness.

By leveraging technology to streamline communication and personalize care, we can pave the way for a healthier future for everyone involved. With ongoing dedication to innovation, the vision of preventive, personalized, and empowering healthcare for all can be realized**.**

REFERENCE

[1] S. Patil, et al., “A Decision Tree Approach for Disease Prediction Using Symptom Data,” International Journal of Computer Applications, vol. 182, no. 23, pp. 1–5, 2020.

[2] R. Kumar and A. Singh, “SVM-Based Classification for Chronic Disease Diagnosis,” Procedia Computer Science, vol. 167, pp. 2060–2069, 2021.

[3] A. Sharma, N. Verma, “Multi-Disease Prediction using Symptom-Based Classification Models,” Journal of Health Informatics in Developing Countries, vol. 16, no. 1, 2022.

[4] P. Gupta and R. Bansal, “A Flask-Based Intelligent System for Medication Recommendation,” IJRTE, vol. 8, no. 4, pp. 321–326, 2019.

[5] V. Reddy, A. Joshi, “Design and Implementation of Web- Based Health Prediction Tool Using Flask,” International Journal of Scientific Research in Computer Science, vol. 9, no. 6, pp. 44–48, 2021.