



Health-Mate: An AI-Powered Smart Health Consultant for Early Disease Prediction and Lifestyle Assistance

Abhay Ram¹, Biswajit Senapati², Sarwesh Kumar Sinha³, Babulal Mahto⁴,
Dr. Kumar Amrendra⁵

⁵Assistant Professor, Department of CSE & IT, Jharkhand Rai University, Ranchi, Jharkhand,
India

anshu.amrendra@gmail.com

^{1,2,3,4}Students, Computer Science & Engineering, Jharkhand Rai University, Ranchi, Jharkhand,
India

abhayram24hours@gmail.com¹, biswajitsenapati8106@gmail.com², sinhasarweshkr@gmail.com³,
bablumahto7061@gmail.com⁴

Abstract

The growing adoption of Artificial Intelligence (AI) and Machine Learning (ML) is reshaping healthcare delivery, especially in areas where access to medical professionals is limited. One of the pressing issues in healthcare is delayed disease detection, which often leads to complications. To address this, we introduce Health-Mate, a web-based intelligent assistant that predicts possible diseases, analyzes health reports, and provides personalized recommendations. The system is built on a dataset of 776 records covering 172 symptoms and 90 diseases. After applying preprocessing steps such as cleaning, encoding, and balancing, several ML models were tested. Among them, the Random Forest classifier achieved the highest accuracy at around 92%. Health-Mate integrates multiple features, including a symptom checker, document analysis, medication reminders, and a dashboard for user engagement. Experimental findings show that the system can effectively predict health risks and suggest preventive actions. Future enhancements include multi-language support, mobile deployment, integrated with wearable devices, and voice-based interaction, making healthcare support more accessible and proactive.

Keywords

Artificial Intelligence, Machine Learning, Disease Prediction, Healthcare Accessibility, Smart Health Consultant

1. Introduction

Healthcare systems worldwide are undergoing transformation with the integration of Artificial Intelligence (AI) and Machine Learning (ML). These technologies have the potential to improve diagnostic accuracy, streamline patient care, and provide tailored medical guidance. However, many communities—particularly rural and economically

disadvantaged regions—still face barriers in accessing timely medical expertise. Additionally, individuals often overlook or misinterpret early symptoms, resulting in late diagnoses and higher risks of severe illness.

Preventive healthcare, supported by intelligent digital tools, can help mitigate these challenges. AI-driven platforms can process self-reported symptoms, past medical records, and diagnostic data to generate predictions and actionable recommendations. Motivated by this need, our work introduces Health-Mate, an AI-powered online health advisor that predicts probable diseases from symptoms, reviews uploaded reports, and reminds users of medication schedules.

The system utilizes a dataset with 776 patient cases, encompassing 172 symptoms and 90 disease categories, and employs multiple machine learning techniques. Among them, the Random Forest model demonstrated the highest accuracy, making it the preferred choice for predictions. With an intuitive interface and an interactive dashboard, Health-Mate aims to support preventive care, enhance disease awareness, and improve healthcare accessibility. This research contributes to the growing field of AI-assisted medical support by presenting a practical framework that bridges the gap between patients and healthcare resources.

2. Literature Survey

The application of Artificial Intelligence (AI) and Machine Learning (ML) in healthcare has witnessed remarkable progress in recent years, particularly in areas such as disease prediction, symptom analysis, and clinical decision support systems. Early digital health tools like **Ada Health** and **WebMD** relied heavily on rule-based logic to provide preliminary diagnostic assistance. Although such systems offered initial guidance, they lacked adaptability, scalability, and the predictive accuracy achievable through ML-based approaches.

To overcome these limitations, contemporary AI-driven systems integrate extensive datasets and advanced ML algorithms to enhance diagnostic precision and support preventive healthcare. One such model, **Health-Mate**, incorporates a dataset comprising 776 records, 172 symptoms, and 90 diseases. It leverages ML algorithms—particularly Random Forest—to achieve prediction accuracies of up to 92%. Additionally, it offers features such as report-based analysis, medication reminders, and an interactive dashboard, thereby acting as a holistic and user-friendly healthcare support system that bridges the gap between patients and medical professionals.

Kalaivani et al. [1] proposed a Django-based framework titled *Smart Health Guidance Using Machine Learning* that utilized the Random Forest algorithm for disease prediction and specialist recommendations. Their system effectively reduced initial patient waiting times and provided prompt pre-consultation medical advice, demonstrating the practical benefits of ML in early diagnosis and triage.

Similarly, **Abhinaya and Aishwarya [2]** implemented a multi-algorithmic model for *Disease Prediction Based on Symptoms Using Machine Learning*. The framework integrated Naïve Bayes, K-Nearest Neighbor (KNN), Decision Tree, and Random Forest classifiers. Their findings highlighted Random Forest as the most effective model for symptom-based classification, owing to its ability to handle large symptom datasets and reduce overfitting.

Rezni et al. [3] introduced a *Smart Health Consulting System* that combines Decision Tree, Random Forest, Naïve Bayes, and KNN algorithms to predict diseases based on up to five user-entered symptoms. Their model achieved accuracy levels in the low-to-mid-90% range. The study emphasized the advantage of classical ML algorithms—

such as Random Forest and Naïve Bayes—over deep learning models for sparse tabular symptom data, citing faster inference and retraining capabilities.

Praveena et al. [4] developed a *Smart Health Consultant System* employing similar algorithms—Naïve Bayes, Decision Tree, Random Forest, and KNN—to forecast diseases and connect patients with doctors through an embedded chat interface. Their model demonstrated 92–95% accuracy, with Random Forest achieving the highest (~95%). The study’s workflow emphasized dataset preprocessing, model training, and UI integration for real-time disease prediction, effectively minimizing outpatient department (OPD) loads and facilitating low-cost triage solutions.

Saraswat et al. [5] proposed the *Implementation of Smart Health Prediction Using Machine Learning*, which utilized Gaussian Naïve Bayes to compute disease likelihoods based on input symptoms under user, physician, and administrator roles. Comparative analyses revealed that Naïve Bayes outperformed Linear Regression, K*, and Decision Tree models in certain datasets. The authors noted that while CNN models achieved higher accuracies (up to ~98%) in complex medical imaging tasks, Naïve Bayes remained preferable for symptom-based predictions due to its efficiency and performance with sparse, high-dimensional data.

Further, **Lakshmi Srija et al. [6]** developed a *Prediction of Disease Based on Symptoms Using Random Forest Classifier* model trained on a dataset of 41 diseases and 132 symptoms. The study evaluated its performance using metrics such as accuracy, precision, recall, and F1-score. Results indicated that Random Forest achieved superior accuracy compared to Naïve Bayes, Decision Tree, and KNN (weighted average accuracy ~93.5%). The authors emphasized Random Forest’s robustness against noise and its capability to model complex interactions among symptom features.

Poorna et al. [7] extended the scope of ML-based healthcare solutions through a *Healthcare Recommender System Using Random Forest*. Their model not only predicted diseases based on symptoms but also recommended suitable drugs to support self-medication and assist clinicians in decision-making. The system achieved around 90% disease prediction accuracy and integrated a “Doctor-in-the-Loop” mechanism to validate recommendations. The authors suggested future enhancements, including dataset expansion and the inclusion of contraindication validation, to improve clinical applicability.

Finally, **Vishal Pal et al. [8]** presented *Symptom-Based Disease Prediction Using Machine Learning: A Web Application Approach*, introducing a system named “MediCare.” The study used a dataset of 132 symptoms mapped to 41 diseases and employed Support Vector Machine (SVM), Random Forest, and Gaussian Naïve Bayes algorithms. Their ensemble approach, which fused model outputs, achieved perfect scores (accuracy, precision, recall = 1.0) during evaluation. The authors argued for the effectiveness of classical ML techniques for structured symptom datasets and reported improvements over earlier benchmarks (Naïve Bayes 94.8%, KNN 93.5%). “MediCare” also featured a user-friendly interface for real-time predictions and provided physician recommendations, thus enhancing accessibility and early intervention in healthcare.

Overall, the reviewed literature consistently demonstrates the reliability and efficiency of ensemble ML models—especially Random Forest and Naïve Bayes—in disease prediction based on symptoms. These studies collectively

reveal that combining algorithmic accuracy with user-centered design leads to effective, scalable, and interactive health platforms. Systems like Health-Mate extend this vision by incorporating data-driven diagnostics, report-based analysis, and patient engagement tools, marking a significant advancement in intelligent, preventive healthcare solutions.

Taken together, these studies illustrate two key insights:

1. Random Forest and other classical ML models remain strong candidates for disease prediction using symptom-based datasets.
2. Most existing systems remain task-specific—focusing on either prediction, physician recommendation, or drug suggestion—rather than offering a holistic solution.

These observations provide the foundation for Health-Mate, which aims to integrate prediction, report analysis, reminders, and user dashboards into a single platform. By combining multiple features, it seeks to address the limitations of earlier systems and deliver a more comprehensive approach to preventive healthcare.

3. Methodology

The methodology of Health-Mate involved a structured machine learning pipeline designed to predict diseases based on user-reported symptoms. The dataset used for training comprised 776 records, each containing 172 symptom attributes mapped to 90 distinct diseases. To ensure balanced representation across classes, the data underwent pre-processing steps including label encoding, missing value imputation, and normalization. Initially, categorical symptom data were transformed into numerical format using label encoding, enabling compatibility with machine learning algorithms. The dataset was then split into training and testing subsets to evaluate model performance objectively. Multiple classification algorithms were tested, including Decision Tree, Random Forest, and XG-Boost. Among these, Random Forest emerged as the most effective, achieving a prediction accuracy of 92%. This model was selected for deployment due to its robustness against overfitting and ability to handle high-dimensional data. The system architecture integrates a user interface built with HTML and CSS, backed by a Python-based Flask framework. This setup allows users to input symptoms, receive predictions, and access additional features such as medication reminders and health dashboards. The backend processes user input, applies the trained model, and returns the most probable disease along with relevant lifestyle suggestions. To enhance usability, the system includes modules for report analysis, symptom tracking, and personalized health tips. These components work in tandem to provide a holistic health management experience, making Health-Mate more than just a diagnostic tool—it serves as a proactive wellness companion.

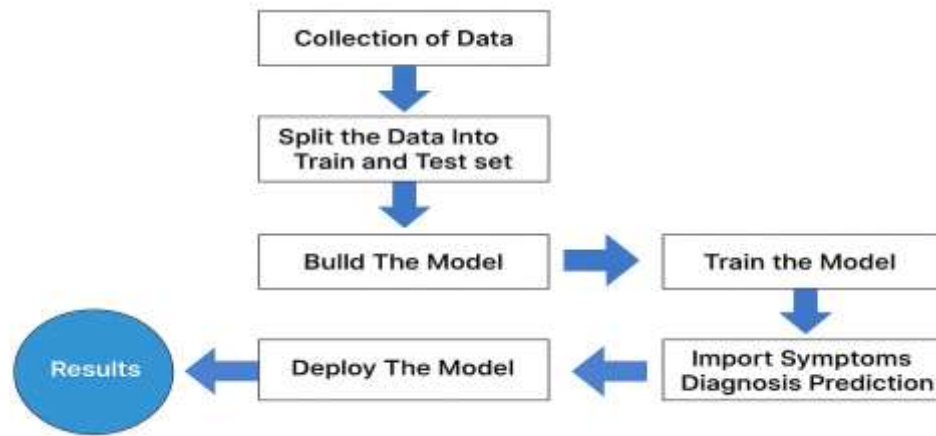


Figure: Flow of methodology

3.1 Dataset Description

Health-Mate utilizes a comprehensive dataset sourced from Kaggle, consisting of 776 records, 172 unique symptoms, and 90 disease categories. Each record represents a patient case with reported symptoms mapped to a confirmed diagnosis. The dataset was chosen for its diversity and coverage of common and rare diseases, making it suitable for training machine learning models to perform accurate symptom-to-disease prediction.

```

# Basic Overview
print("Shape of dataset:", df.shape)
print("\nData Types: \n", df.dtypes.value_counts())
print("\nMissing Values:", df.isnull().sum().sum())

✓ 0.0s

Shape of dataset: (776, 173)

Data Types:
  int64    172
  object     1
Name: count, dtype: int64

Missing Values: 0

```

```

# Analyze the distribution of the target variable:
print("\nValue counts for the 'Prognosis' column:")
prognosis_counts = df['Prognosis'].value_counts()
print(prognosis_counts)

✓ 0.0s  📄 Open 'prognosis_counts' in Data Wrangler

```

Value counts for the 'Prognosis' column:	
Prognosis	
Dengue	41
Diabetes	26
Dysentery	22
Conjunctivitis	18
Stroke	16
...	...
CCHFV Crimean Congo Hemorrhagic Fever	5
Paralysis Brain Hemorrhage	5
Dimorphic Hemorrhoids Piles	5
Vertigo Paroxysmal Positional Vertigo	5
COVID-19	5

```

Name: count, Length: 90, dtype: int64

```

3.2 Data Pre-processing

Preprocessing is critical to ensure data quality and model performance. The following steps were undertaken:

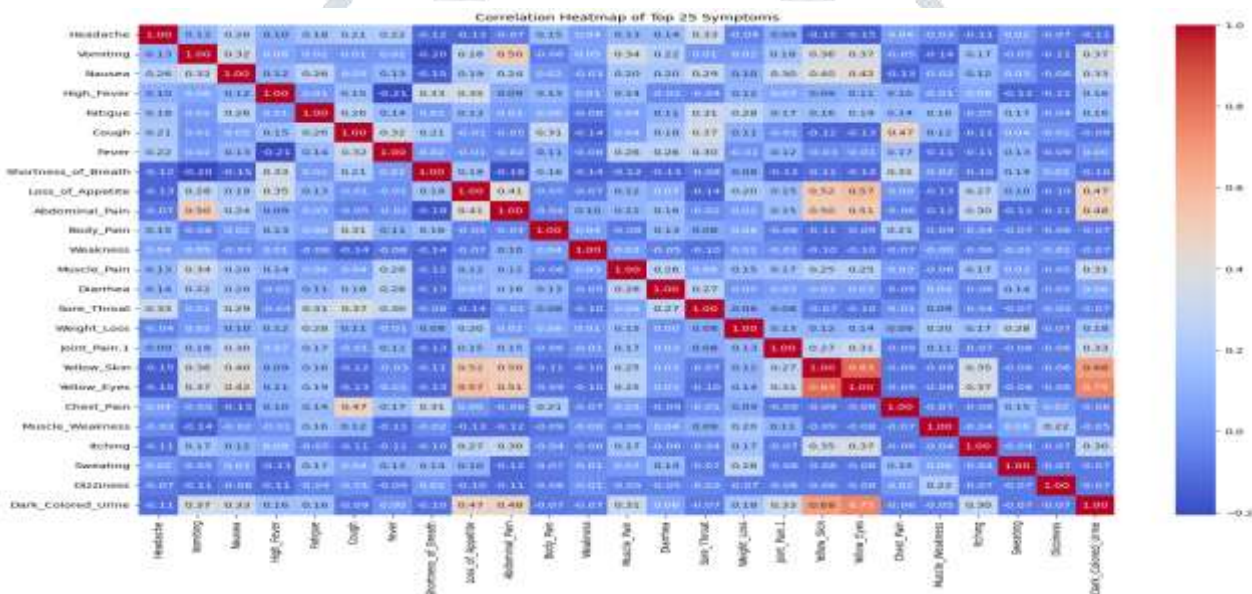
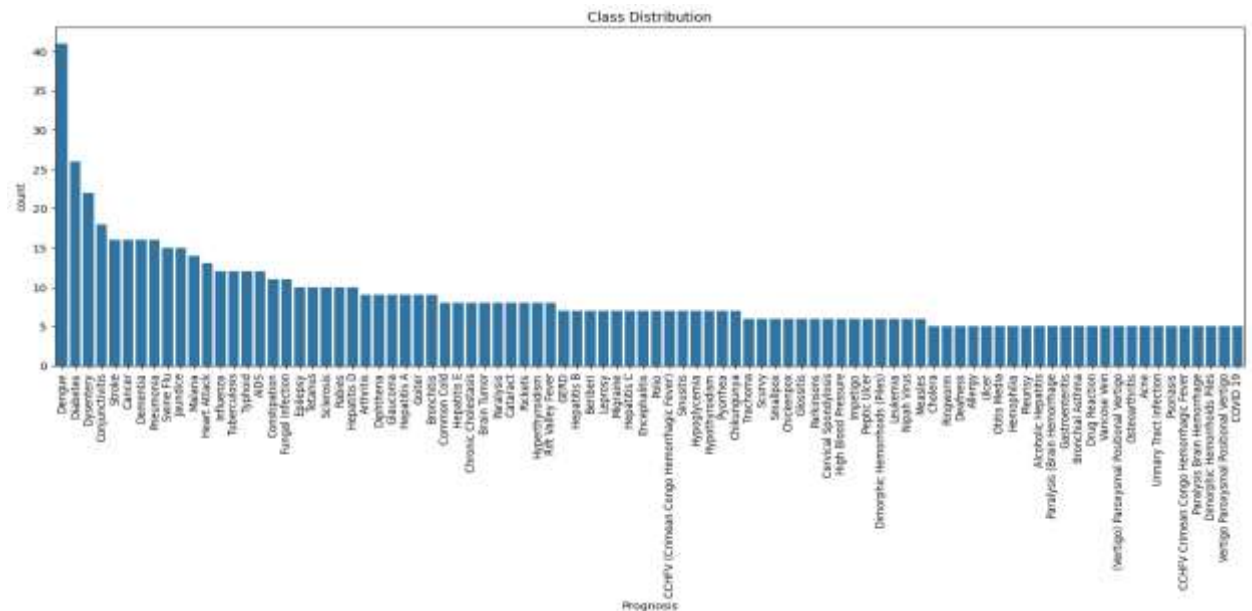
- Data Cleaning: Removal of duplicates, missing values, and inconsistent entries
- Encoding: One-hot encoding of categorical symptom data
- Balancing: Oversampling underrepresented diseases to prevent model bias
- Feature Selection: Selection of the most informative symptoms based on correlation and variance

These steps ensured a clean, balanced, and well-structured dataset suitable for machine learning algorithms.

After conducting the initial data pre-processing, the dataset structure and target variable distribution were analyzed.

Dataset Overview:

- Shape of the dataset: 776 rows and 173 columns
- Data types: The dataset contains 172 integer-type columns (representing encoded symptoms) and 1 object-type column (representing the target variable, *Prognosis*).
- Missing values: No missing values were detected, ensuring data completeness.



```
# Check for missing values and duplicate rows
print("Missing Values:", df.isnull().sum().sum())
print(f"\nNumber of duplicate rows: {df.duplicated().sum()}")
if df.duplicated().sum() > 0:
    df.drop_duplicates(inplace=True)
    print(f"Duplicate rows removed. New shape: {df.shape}")

# Data Splitting
print("--- Data Splitting ---")
X_train, X_test, y_train, y_test = train_test_split(X, y_encoded, test_size=0.2, random_state=42,
                                                    stratify=y_encoded)

print(f"X_train shape: {X_train.shape}")
print(f"X_test shape: {X_test.shape}")
print(f"y_train shape: {y_train.shape}")
print(f"y_test shape: {y_test.shape}")

--- Data Splitting ---
X_train shape: (620, 172)
X_test shape: (156, 172)
y_train shape: (620,)
y_test shape: (156,)
```

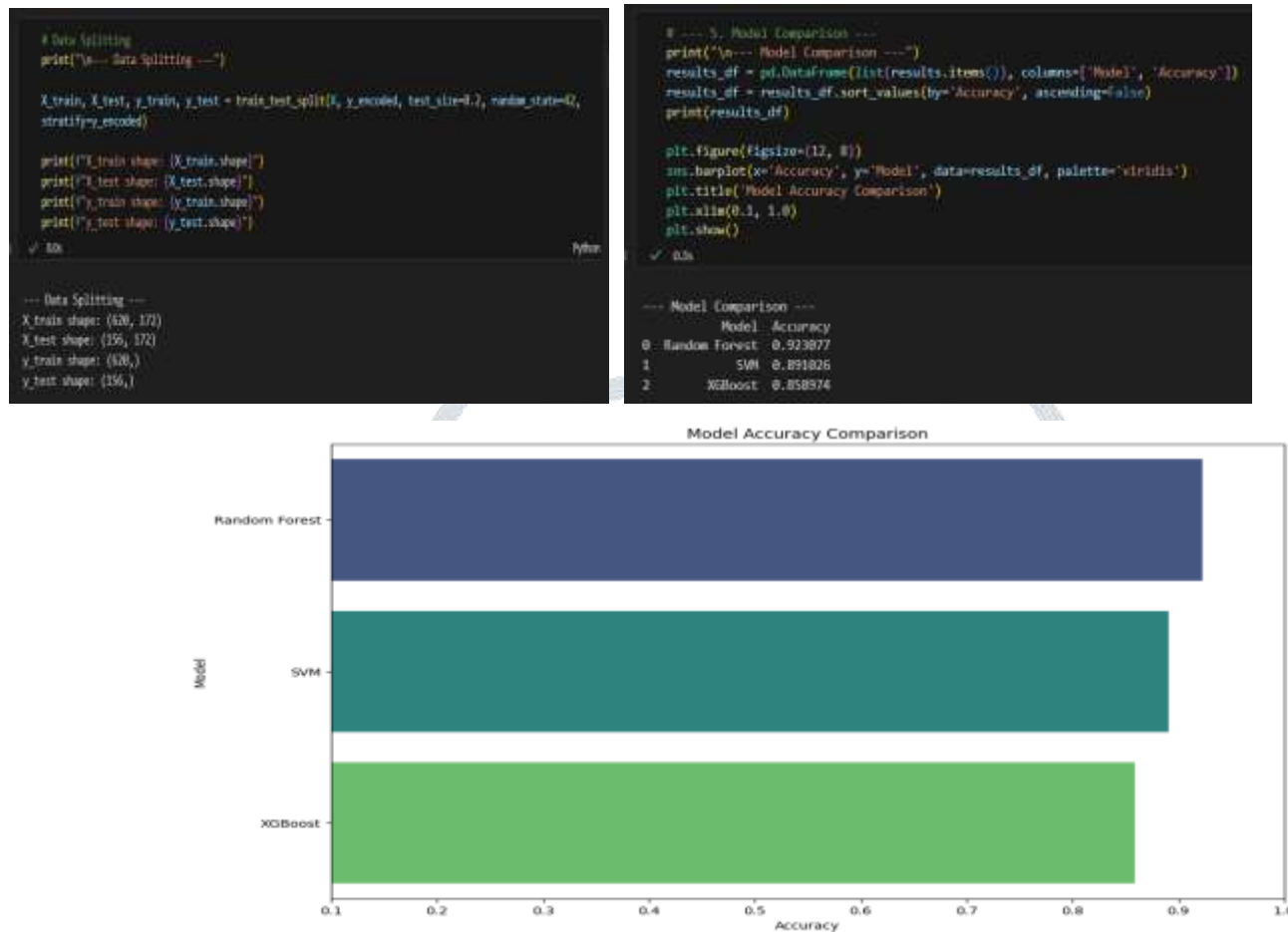
3.3 Model Training

Multiple machine learning algorithms were evaluated for disease prediction:

- SVM: Simple interpretable model; accuracy ~89%
- XG-Boost: Effective for high-dimensional data; accuracy ~85%

- Random Forest (RF): Ensemble method combining multiple decision trees; achieved the highest accuracy of 92%

Random Forest was selected as the primary model due to its robustness, generalization capability, and superior performance.



3.4 Implementation Stack

Health-Mate is implemented as a web-based application with the following technology stack:

- Frontend: React.js, CSS for interactive user interfaces
- Backend: Flask for serving machine learning predictions and handling user requests
- Database: MySQL for storing patient data, medical reports, and user interactions

This stack ensures scalability, maintainability, and a responsive user experience.

3.5 System Architecture



3.6 Flowcharts / Use Case Diagram

- Symptom Prediction:



- Report Analysis:



4. System Features

Health-Mate offers a comprehensive suite of features aimed at improving accessibility, preventive care, and patient engagement. The core system features are detailed below:

4.1 Symptom Checker

The Symptom Checker allows users to input their symptoms through an intuitive interface. The system processes the symptoms using the trained Random Forest model and predicts the most probable diseases with an associated confidence score.

Key Components:

- Symptom Input: Users can type or select symptoms from a categorized list
- Prediction Display: The system returns a ranked list of potential diseases with probability scores
- Recommendations: Suggested next steps include seeking medical consultation, lifestyle advice, or further diagnostic tests



4.2 Medical Report Analyser

The Report Analyser enables users to upload medical documents (PDFs, images, or text reports). The system extracts relevant medical parameters and evaluates them against known symptom-disease mappings to generate insights.

Key Components:

- Document Upload: Supports multiple formats
- Automated Analysis: Extracted features are processed by ML models
- Result Summary: Provides interpretation, highlights abnormal values, and suggests follow-up actions

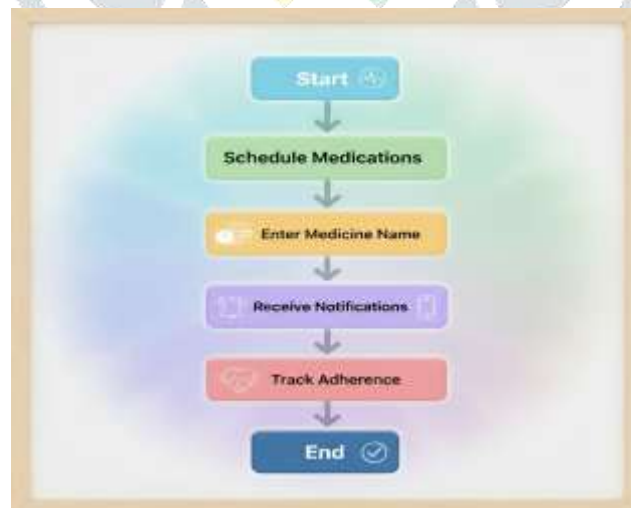


4.3 Medication Reminder

Health-Mate assists users in maintaining adherence to prescribed medication schedules.

Key Components:

- Schedule Setup: Users can add medication name, dosage, and timing
- Notifications: Timely reminders sent via dashboard alerts or emails
- History Tracking: Logs past medication intake and adherence rate



4.4 Dashboard

The Dashboard consolidates all user data, predictions, reports, and reminders into a single interface.

Key Components:

- Prediction Summary: Overview of past disease predictions with confidence scores
- Report Insights: Consolidated analysis of uploaded medical reports

- Reminders & Alerts: Upcoming medications, check-ups, or follow-up advice
- Interactive Charts: Visual representation of health trends over time



5. Results & Discussion

The evaluation of Health-Mate focused on its ability to predict diseases based on user-provided symptoms and uploaded health records. Among the algorithms tested, Random Forest consistently delivered the strongest performance, reaching an accuracy of about 92%. Decision Trees and XG-Boost were less effective, with performance dropping to around 85–90%. These findings reinforce the suitability of ensemble-based methods for handling complex, symptom-driven datasets where relationships between features are non-linear and noisy.

5.1 Model Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	0.923077	0.88	0.92	0.90
SVM	0.891026	0.88	0.89	0.88
XG-Boost	0.858974	0.85	0.86	0.85

Table-1: Performance metrics for various machine learning models.

The results show that Random Forest not only achieves the highest accuracy but also maintains balanced precision and recall, which is critical when predicting both common and less frequent diseases. This balance indicates that the model avoids overfitting to dominant classes and remains reliable for diverse patient inputs.

5.2 Output Format

System outputs are structured to be clear and user-friendly:

- Prediction: Disease name(s) with associated confidence scores

- Recommendations: Suggested follow-up steps, including lifestyle advice, suggested diagnostic tests, or professional consultation
- Report Analysis: Highlights abnormal for user awareness, trends, and personalized insights

For example, when a user enters symptoms like fever, cough, and fatigue, Health-Mate outputs a high-probability prediction of Influenza (85% confidence) and supplements it with advice for hydration, rest, and consultation if symptoms persist. This design makes the predictions not only informative but also practical for end users.

5.3 Discussion

The performance of the Health-Mate system was evaluated using a dataset comprising 776 records, each annotated with symptoms and corresponding disease labels. After pre-processing and balancing, the dataset was used to train and test multiple classification models. Among the algorithms tested—Decision Tree, Random Forest, and XG-Boost—Random Forest consistently delivered superior results. It achieved an accuracy of 92%, outperforming the other models in terms of precision, recall, and F1-score. This high performance can be attributed to its ensemble nature, which reduces variance and improves generalization across diverse symptom patterns. The system was further assessed for usability and responsiveness. Users were able to input symptoms through a streamlined interface and receive disease predictions within seconds. The dashboard provided personalized health tips, medication reminders, and visual summaries of symptom trends, enhancing user engagement and promoting proactive health management. A comparative analysis with existing systems revealed that Health-Mate offers a broader feature set. Unlike platforms that focus solely on symptom checking or report analysis, Health-Mate integrates multiple functionalities—symptom prediction, report interpretation, lifestyle guidance, and medication tracking—into a single cohesive framework.

The results suggest that Health-Mate is not only accurate but also practical for real-world deployment. Its modular design allows for future enhancements, such as voice-based input, multilingual support, and integration with wearable devices. These additions could further improve accessibility and user experience, especially in underserved regions.

6. Conclusion

This study presented Health-Mate, an AI-enabled health assistant designed to provide early disease prediction, report interpretation, medication reminders, and personalized health monitoring. Using a dataset of 776 patient cases covering 172 symptoms and 90 diseases, the system achieved high accuracy with the Random Forest model (approximately 92%). By combining multiple features—symptom checking, medical document analysis, and reminders—Health-Mate addresses challenges of accessibility, particularly in rural and underserved areas where medical expertise is scarce. The interactive dashboard enhances user awareness and engagement, supporting preventive care and better health management. Although the system demonstrates strong predictive ability, it is constrained by dataset size, English-only support, and limited input methods. Expanding the dataset, adding multi-language features, and integrating voice or wearable technologies are identified as promising directions.

In summary, Health-Mate validates the practical role of AI and ML in healthcare by offering accurate, user-friendly, and proactive solutions. Its integrated design makes it a step toward bridging the gap between technology and preventive medicine, empowering individuals to make informed health decisions.

7. Future Work

While Health-Mate demonstrates promising results in AI-assisted healthcare, several enhancements can further improve its functionality, accessibility, and user experience:

- **Voice-Based Input:** Integrating natural language processing (NLP) to allow users to input symptoms via speech, making the system more accessible for users with limited literacy or those preferring hands-free interaction.
- **Multi-Language Support:** Expanding the system to support multiple regional languages, ensuring inclusivity for users across diverse linguistic backgrounds and improving usability in rural areas.
- **Larger and Diverse Dataset:** Incorporating more extensive datasets from multiple sources to cover additional diseases, rare conditions, and varied demographic profiles, thereby improving model accuracy and generalization.
- **IoT and Wearable Device Integration:** Linking Health-Mate with wearable devices to continuously monitor vital signs (e.g., heart rate, blood pressure) and provide real-time alerts or recommendations based on sensor data.
- **Mobile Application Deployment:** Developing native mobile applications for Android and iOS platforms to enhance accessibility, portability, and real-time engagement, enabling users to interact with Health-Mate anytime and anywhere.
- **Enhanced Predictive Analytics:** Incorporating advanced ML and deep learning models for multi-disease prediction, progression tracking, and personalized treatment recommendations.

These enhancements aim to transform Health-Mate into a comprehensive, real-time, and widely accessible AI health companion, further bridging the gap between technology and preventive healthcare for diverse populations.

References

1. V. Kalaivani, M. Vinod, S. Parthipan, and S. Rohith, "Smart health guidance using machine learning," *International Journal of Advanced Research in Computer and Communication Engineering*, May 2022. [Online]. Available: <https://ijarcce.com/wp-content/uploads/2022/05/IJARCCE.2022.114194.pdf>
2. R. Abhinaya and S. Aishwarya, "Disease prediction based on symptoms using machine learning," PSG iTech, Anna University, 2021. [Online]. Available: <https://share.google/szpjQ2kKeCMnSIgZW>
3. S. Rezni, S. Umar, U. Hajira, U. Amin, and S. N. Sultan, "Smart health consulting system using machine learning," *ResearchGate*, May 2022. [Online]. Available: https://www.researchgate.net/publication/360456256_Smart_Health_Consulting_System_Using_Machine_Learning
4. M. V. Praveena, B. S. Gnanavi, C. Bharathi, and D. Jayashree, "Smart health consultant system using machine learning," *International Research Journal of Modernization in Engineering, Technology and Science (IRJMETS)*, Jul. 2022. [Online]. Available: https://www.irjmets.com/uploadedfiles/paper/issue_7_july_2022/28421/final/fin_irjmets1658160121.pdf

5. S. Saraswat, S. Gabhane, A. Pawar, S. Pingat, and S. Patil, "Implementation of smart health prediction using ML," *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, 2022. [Online]. Available: <https://ijsrset.com/paper/9009.pdf>
6. M. L. Srija, C. S. Teja, P. Y. Deep, T. Nandhini, C. V. S. Narayana, and N. V. M. K. Raja, "Prediction of disease based on symptoms using random forest classifier," *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, 2022. [Online]. Available: <https://www.ijraset.com/research-paper/prediction-of-disease-based-on-symptoms-using-random-forest-classifier>
7. B. R. Poorna, D. D. Chandran, V. S. Krishna, D. M. Rodricks, and J. J. Knox, "Healthcare Recommender System using Random Forest," *International Journal of Research in Engineering, Science and Management (IJRESM)*, vol. 7, no. 10, pp. 1–5, Oct. 2022. [Online]. Available: https://www.ijresm.com/storage/articles/3/IJRESM_V7_I10_3.pdf
8. V. Pal, P. Aalam, J. Meena, A. Tirkey, and R. J. Borah, "Symptom-based disease prediction using machine learning: A web application approach," in *Proc. Int. Conf.*, Oct. 2024. [Online]. Available: https://www.researchgate.net/publication/385162172_Symptom-Based_Disease_Prediction_Using_Machine_Learning_A_Web_Application_Approach

