



Classification and Analysis of Disease over Symptoms using AI

¹Peetambar, ²Amir Ansari, ³Samrat Kartikey Maurya, ⁴Mr. Dileep Kumar Gupta

Scholar B. Tech Final Year, Department of Computer Science & Engineering, Goel Institute of Technology & Management, Lucknow, Uttar Pradesh, India¹⁻³

Assistant Professor, Department of Computer Science & Engineering, Goel Institute of Technology & Management, Lucknow, Uttar Pradesh, India⁴

Abstract: This paper presents a smart healthcare system that integrates machine learning-based disease prediction with an online appointment booking platform. The system aims to address the problem of delayed diagnosis and limited access to medical consultation by combining predictive analytics with real-time healthcare services. Machine learning algorithms such as Support Vector Machine (SVM) and Logistic Regression are used to predict diseases like diabetes, heart disease, and Parkinson's based on user input data.

In addition to prediction, the system provides a web-based interface where patients can easily browse doctors and book appointments. The platform includes role-based dashboards for patients, doctors, and administrators, enabling efficient management of appointments, patient records, and doctor availability. The integration of prediction and booking ensures that users can take immediate action after receiving health insights.

The proposed system improves early detection, reduces manual effort, and enhances accessibility to healthcare services. It also provides a scalable and user-friendly solution that can be deployed in real-world healthcare environments. This approach demonstrates how the combination of machine learning and full-stack development can significantly improve healthcare delivery and patient outcomes.

The system is designed to be user-friendly, scalable, and efficient for real-world healthcare applications. It ensures secure data handling and smooth integration between prediction and booking modules. The approach reduces the gap between diagnosis and consultation, improves patient engagement, and helps healthcare providers manage patient flow effectively.

Keyword: Artificial Intelligence, Machine Learning, Disease Prediction, Healthcare Analytics, Diabetes Prediction, Heart Disease Detection, Parkinson's Prediction, Predictive Modeling, Clinical Decision Support System, Streamlit Web Application, Medical Data Analysis, Early Diagnosis, Health Monitoring System, AI in Healthcare, Appointment Booking System Integration

INTRODUCTION

In recent years, the integration of Artificial Intelligence (AI) and Machine Learning (ML) in healthcare has significantly transformed the way diseases are diagnosed and managed. Early detection of chronic diseases such as diabetes, heart disease, and Parkinson's disease plays a crucial role in improving patient outcomes and reducing healthcare costs. However, traditional diagnostic methods often require extensive clinical tests, expert consultation, and significant time, which may delay timely treatment.

To address these challenges, this project proposes an AI-powered web-based system named **Classification and Analysis of Disease over Symptoms using AI**, which aims to provide early disease prediction using machine learning techniques. The system is designed to analyze user-provided medical parameters and predict the likelihood of multiple diseases, including diabetes, heart disease, and Parkinson's disease. By leveraging trained machine learning models, the system assists users in identifying potential health risks at an early stage.

The proposed system utilizes pre-trained models that are loaded and executed using a user-friendly interface built with Streamlit. Users can input relevant medical data such as glucose levels, blood pressure, cholesterol levels, and other clinical attributes. Based on this input, the system processes the data and generates predictions indicating whether a person is likely to have a particular disease. This approach not only reduces dependency on manual diagnosis but also



provides instant results, making it highly efficient and accessible.

In addition to disease prediction, the system integrates an appointment booking feature that allows users to consult specialized doctors after receiving their results. This integration bridges the gap between prediction and professional medical advice, ensuring that users can take appropriate actions based on the system's output. The combination of predictive analytics and healthcare services enhances the overall effectiveness of the platform.

In conclusion, the **Classification and Analysis of Disease over Symptoms using AI** system demonstrates how modern technologies such as machine learning and web applications can be effectively utilized to improve healthcare services. It not only empowers individuals to monitor their health but also contribute to the advancement of intelligent healthcare systems.

BACKGROUND

The healthcare industry has witnessed rapid advancements with the integration of digital technologies, particularly Artificial Intelligence (AI) and Machine Learning (ML). These technologies have enabled the development of intelligent systems capable of analyzing large volumes of medical data and providing accurate predictions. Chronic diseases such as diabetes, heart disease, and Parkinson's disease are among the leading causes of mortality worldwide. Early detection of these diseases is essential to prevent complications and improve patient survival rates.

Traditionally, disease diagnosis relies heavily on clinical expertise, laboratory tests, and manual evaluation of patient data. While these methods are effective, they are often time-consuming, costly, and may not be easily accessible in remote or underdeveloped regions. Additionally, the increasing patient load in healthcare facilities creates a need for automated systems that can assist healthcare professionals in making quicker and more accurate decisions.

Machine Learning algorithms have shown great potential in addressing these challenges by identifying patterns and relationships within medical datasets. Supervised learning techniques, in particular, are widely used in disease prediction, where models are trained on historical patient data to classify or predict health conditions. These models can process multiple input parameters such as blood pressure, glucose levels, heart rate, and other clinical indicators to determine the probability of a disease.

IDENTIFICATION OF THE ISSUE

The increasing prevalence of chronic diseases such as diabetes, heart disease, and Parkinson's disease has become a major concern in modern healthcare systems. One of the primary issues is the lack of early detection, as many individuals remain unaware of their health conditions until symptoms become severe. Traditional diagnostic processes are often time-consuming, expensive, and dependent on specialized medical professionals, which limits accessibility, especially in rural and underdeveloped areas.

Another significant issue is the absence of integrated digital platforms that can provide both disease prediction and immediate medical consultation. Existing systems typically focus on a single disease and lack user-friendly interfaces, making them difficult for non-technical users to operate.

Additionally, delays in diagnosis can lead to complications, increased treatment costs, and higher mortality rates.

Another significant issue is the absence of integrated digital platforms that can provide both disease prediction and immediate medical consultation. Existing systems typically focus on a single disease and lack user-friendly interfaces, making them difficult for non-technical users to operate. Additionally, delays in diagnosis can lead to complications, increased treatment costs, and higher mortality rates.

Therefore, there is a need for an efficient, accessible, and intelligent system that can provide early disease prediction along with easy access to healthcare services, ensuring timely intervention and better patient outcomes.

DRAWBACK OF THE CURRENT SYSTEM

The current healthcare diagnostic systems face several limitations that reduce their efficiency and accessibility. One major drawback is the dependency on manual diagnosis, which requires physical visits to hospitals and consultation with medical experts. This process is time-consuming and may delay early detection of diseases, especially in critical cases.



Additionally, diagnostic procedures often involve multiple laboratory tests, increasing the overall cost of healthcare.

Another limitation is that most existing digital systems are designed to detect only a single disease, lacking a unified platform for multi-disease prediction. This creates inconvenience for users who need to undergo different tests and use multiple applications for different health conditions. Moreover, many systems do not provide real-time results, reducing their effectiveness in emergency situations.

The lack of integration with healthcare services, such as doctor consultation or appointment booking, further limits their usefulness. As a result, users are unable to take immediate action after diagnosis, leading to delays in treatment and poorer health outcomes.

PROPOSED FRAMEWORK

The proposed framework presents an AI-based multi-disease prediction system named **Classification and Analysis of Disease over Symptoms using AI**, designed to provide early detection of diabetes, heart disease, and Parkinson's disease through a unified web platform. The system utilizes pre-trained machine learning models that analyze user-input medical parameters and generate instant predictions.

The system is implemented using Streamlit, ensuring accessibility and ease of use. Additionally, it integrates an appointment booking feature, allowing users to consult medical professionals immediately after receiving predictions. This framework enhances early diagnosis, reduces dependency on traditional methods, and provides a fast, reliable, and accessible healthcare solution.

SUGGESTED SYSTEM ARCHITECTURE

The suggested system architecture of **Classification and Analysis of Disease over Symptoms using AI** is based on a three-layer design: user interface, processing layer, and prediction module. The user interface is developed using Streamlit, where users input medical parameters through interactive forms. This layer ensures a simple and user-friendly experience.

The processing layer handles data validation and conversion of user inputs into a suitable format for model prediction. It acts as a bridge between the interface and the machine learning models.

The prediction module consists of pre-trained machine learning models for diabetes, heart disease, and Parkinson's disease. These models analyze the input data and generate prediction results. Additionally, the system integrates an external appointment booking service, enabling users to consult doctors after receiving results.

The suggested system architecture of **Classification and Analysis of Disease over Symptoms using AI** is designed as a component-based healthcare prediction system. The user interacts with the system through a web interface developed using Streamlit. This interface allows users to navigate the home page, select a disease prediction option, enter medical parameters, and view prediction results.

The component diagram shows the internal connection between user, web interface, input module, preprocessing module, prediction engine, model repository, database, and doctor appointment service. It also represents access control, data flow, model loading, and appointment integration. This architecture ensures fast prediction organized processing, and better medical consultation support.

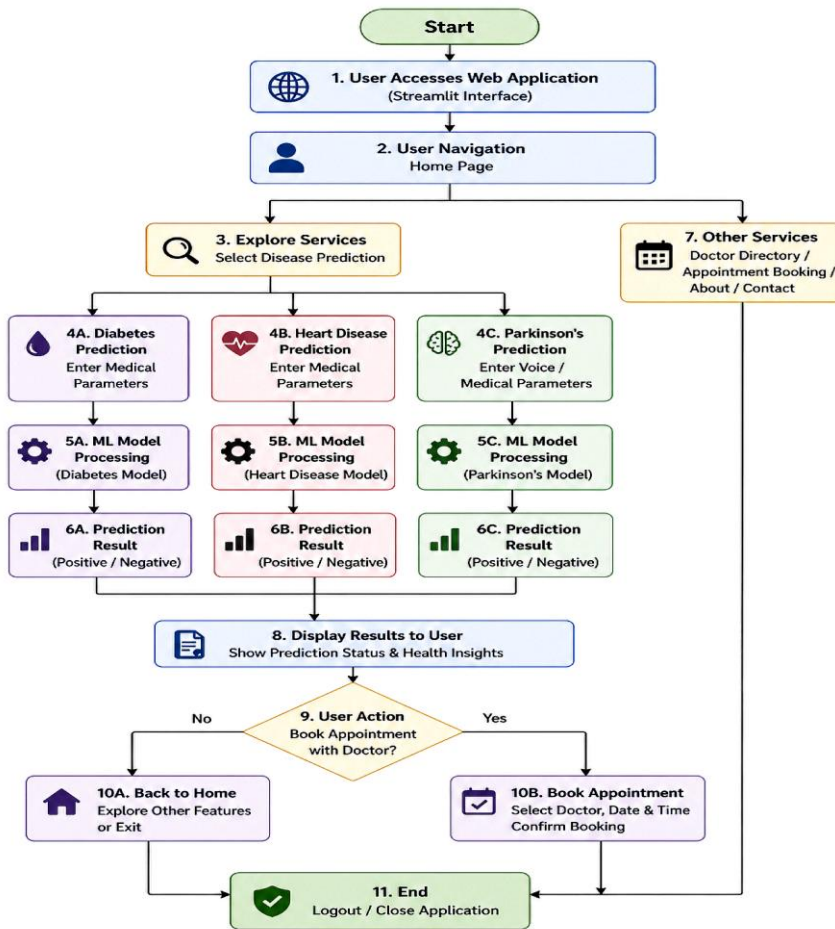


Figure 1: The System Architecture

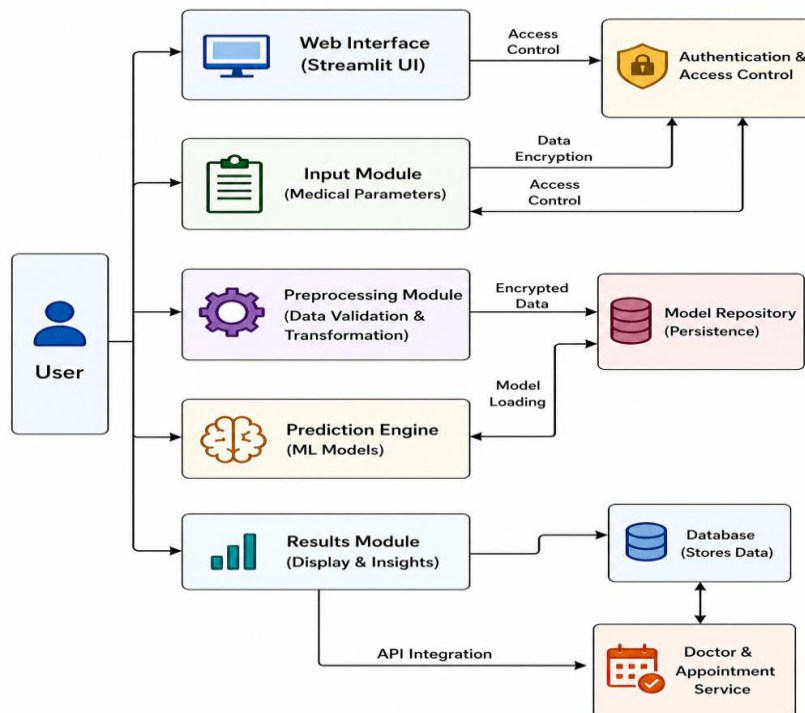


Figure 2: The Components Diagram



The suggested system architecture of **Classification and Analysis of Disease over Symptoms using AI** is designed as a modular and efficient system that integrates machine learning with a web-based healthcare application. The system starts with a disease dataset, which is used to train predictive models. The data first goes through preprocessing, where missing values are handled, noise is reduced, and data is normalized. After preprocessing, feature selection techniques are applied to identify the most relevant attributes for accurate prediction.

The processed dataset is then divided into training and testing sets using a standard training-testing split method. Machine learning algorithms are trained on the training data to develop a predictive model. This model is evaluated using testing data to ensure performance and accuracy. The final trained model is capable of classifying users as healthy or disease-affected based on input parameters.

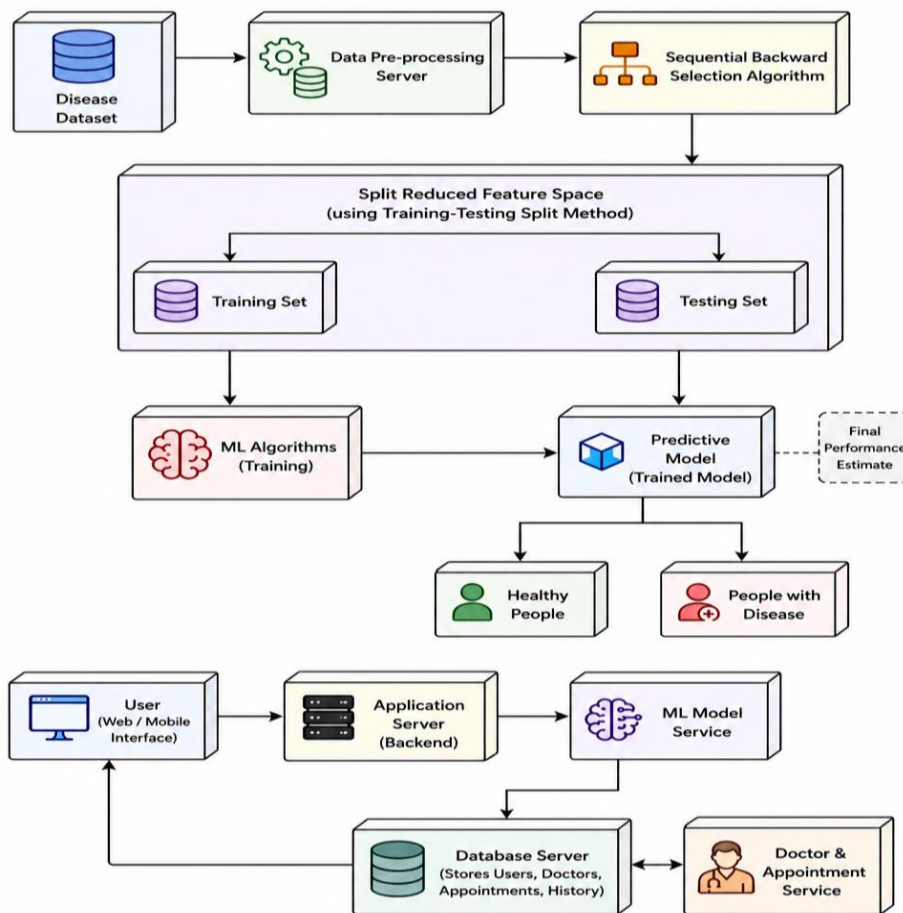


Figure 3: The Deployment Diagram

IMPLEMENTATION

The implementation of **Classification and Analysis of Disease over Symptoms using AI** is carried out using a combination of machine learning and web technologies to develop a functional multi-disease prediction system. The application is built using Python, with Streamlit used to create an interactive and user-friendly web interface. Separate machine learning models are developed for diabetes, heart disease, and Parkinson's disease using preprocessed medical datasets.

The frontend allows users to navigate between different disease prediction modules and enter required parameters. The system then displays the prediction result instantly in a clear format. Additionally, the application integrates a doctor and appointment booking system, where user and doctor details are managed through a database.

This implementation ensures real-time prediction, easy accessibility, and seamless integration between AI models and



healthcare services

7.1 Parkinsons Disease Prediction

Under the implementation phase, the Parkinson's disease prediction module is developed using a trained machine learning model. It analyzes voice-related features such as jitter, shimmer, and frequency values provided by the user to predict the presence of Parkinson's disease accurately.

7.1.1 Attribute Information

In this project, the Parkinson's disease prediction model uses multiple **voice-based biomedical features** as input. These attributes help in identifying irregularities in speech patterns, which are common indicators of Parkinson's disease.

1. Frequency Features:

- **MDVP: Fo (Hz):** Average fundamental frequency of voice.
- **MDVP: Fhi (Hz):** Maximum vocal frequency.
- **MDVP: Flo (Hz):** Minimum vocal frequency.
These values represent the pitch range of the voice.

2. Jitter Metrics (Frequency Variation):

- **Jitter (%), Jitter (Abs), RAP, PPQ, DDP**
These features measure small variations in frequency, indicating instability in vocal fold vibrations.

3. Shimmer Metrics (Amplitude Variation):

- **Shimmer, Shimmer (dB), APQ3, APQ5, APQ, DDA**
These represent variations in amplitude (loudness) of the voice signal.

4. Noise Measurements:

- **NHR (Noise-to-Harmonics Ratio) and HNR (Harmonics-to-Noise Ratio)**
- These determine the level of noise present in the voice signal.

5. Nonlinear Features:

- **RPDE, DFA, D2**
These analyze the complexity and irregular patterns in speech signals.

6. Frequency Spread Metrics:

- a. **spread1 and spread2**
These indicate variations in fundamental frequency distribution.

7. PPE (Pitch Period Entropy):

Measures randomness in pitch patterns.

Status (Target Variable):

Indicates output as **0 (Healthy)** or **1 (Parkinson's)**.

7.1.2 Comparison of Models

Sr. No.	Model Name	Train Accuracy (%)	Test Accuracy (%)	AUC Score
0	Logistic Regression	82.524272	74.576271	0.887879
1	Decision Tree Classifier	83.980583	88.135593	0.910606
2	AdaBoost	83.980583	88.135593	0.854545
3	Random Forest Classifier	99.029126	84.745763	0.841667
4	kNN	100.000000	98.305085	0.966667
5	SVM	100.000000	94.915254	0.992424
6	XGBoost	100.000000	91.525424	0.956061



We can say that the KNN model performs well for our dataset, but SVM provides higher AUC, indicating better classification performance, improved generalization, and more reliable predictions across different data samples overall.

7.1.3 Classification Report

Class	Precision	Recall	F1-Score	Support
0 (Healthy)	1.00	0.93	0.97	15
1 (Parkinson)	0.98	1.00	0.99	44
Accuracy			0.98	59
Macro Avg	0.99	0.97	0.98	59
Weighted Avg	0.98	0.98	0.98	59

In this project, the Parkinson's disease prediction model shows strong performance with approximately 98% testing accuracy. The classification report presents high precision, recall, and F1-score for both healthy and Parkinson's classes, indicating that the trained model can effectively classify users based on voice-related biomedical input features.

7.2 Diabetes Disease Prediction

The prediction employs several supervised machine learning approaches to ascertain a patient's probability of early diabetes onset. It utilizes data on the preferences of affected and unaffected individuals to ascertain if a person is afflicted by a certain disease.

7.2.1 Attribute Information

1. Pregnancies
2. Glucose
3. Blood pressure
4. Skin Thickness
5. Insulin
6. BMI
7. Diabetes Pedigree Function
8. Age

7.2.2 Comparison model

```
# Accuracy on test set
print("Logistic Regression:", str(acc_logreg * 100))
print("K Nearest Neighbors:", str(acc_knn * 100))
print("Support Vector Classifier:", str(acc_svm * 100))
print("Naive Bayes:", str(acc_nb * 100))
print("Decision Tree:", str(acc_dt * 100))
print("Random Forest:", str(acc_rf * 100))
```

```
Logistic Regression: 74.576271
K Nearest Neighbors: 98.305085
Support Vector Classifier: 94.915254
Naive Bayes: 71.428571
Decision Tree: 88.135593
Random Forest: 84.745763
```



7.2.3 Classification Report

```
# Classification report
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred_knn))
```

	precision	recall	f1-score	support
0 (Healthy)	1.00	0.93	0.97	15
1 (Parkinson)	0.98	1.00	0.99	44
accuracy			0.98	59
macro avg	0.99	0.97	0.98	59
weighted avg	0.98	0.98	0.98	59

7.3 Heart Disease Prediction

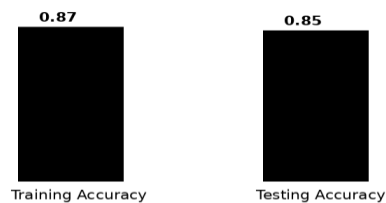
In this project, the heart disease prediction module uses medical data of both affected and healthy individuals to determine the patient's condition. The system applies a supervised machine learning technique, specifically **Logistic Regression**, to analyze user-provided health parameters and generate accurate prediction results.

7.3.1 Attribute Information

- 1.Age
2. Sexual
- 3.Types of Chest Pain
- 4.Blood pressure at rest
- 5.Cholesterol serum
- 6.Blood sugar fasting
- 7.Resting Heart Diagram Outcome
- 8.Peak heart rate attained
- 9.Workout Lessened Angina
10. Fluorescence-colored vessels

7.3.2 Accuracy Results

7.3.2 Accuracy Results (Heart Disease)



7.4 Formulated Algorithm

Step 1: Initially, the system loads all the pre-trained machine learning models required for prediction. These models are stored in .sav format and are loaded into memory at runtime, eliminating the need for retraining and ensuring faster



execution.

Step 2: The user interacts with the Streamlit-based web interface and selects the desired disease prediction module such as Diabetes, Heart Disease, or Parkinson’s Disease. The system then prompts the user to enter relevant medical parameters.

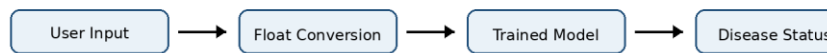
Step 3: The entered input values are collected and converted into numerical (float) format. The system ensures that the data is structured in the same order and format as used during model training to maintain prediction accuracy.

Step 4: The processed input data is passed to the corresponding trained machine learning model using the `predict()` function. The model evaluates the input features and classifies the data into binary output indicating disease presence (1) or absence (0).

7.4 Formulated Algorithm - Model Usage Chart

Disease Module	Input Features	Algorithm Used	Prediction Output
Diabetes Prediction	8 Medical Features	Linear SVM	Positive / Negative
Heart Disease Prediction	13 Medical Features	Logistic Regression	Positive / Negative
Parkinson’s Prediction	22 Voice-Based Features	Linear SVM	Positive / Negative

Algorithm Flow:



Used Algorithms: Linear SVM for Diabetes and Parkinson’s; Logistic Regression for Heart Disease.

OUTCOME

The output of the proposed **Classification and Analysis of Disease over Symptoms using AI** system is presented through an interactive Streamlit-based user interface. After the user enters the required medical parameters and clicks the prediction button, the system processes the input data and passes it to the corresponding trained machine learning model. The model then generates a binary result indicating whether the disease is present or not present.

The result is displayed in a clear textual format such as “Positive” or “Negative,” making it easy for users to understand their health condition. Each disease module (Diabetes, Heart Disease, and Parkinson’s) provides separate outputs based on the selected input parameters. The system ensures real-time response, allowing users to perform multiple predictions efficiently.



8.1 Diabetes Prediction

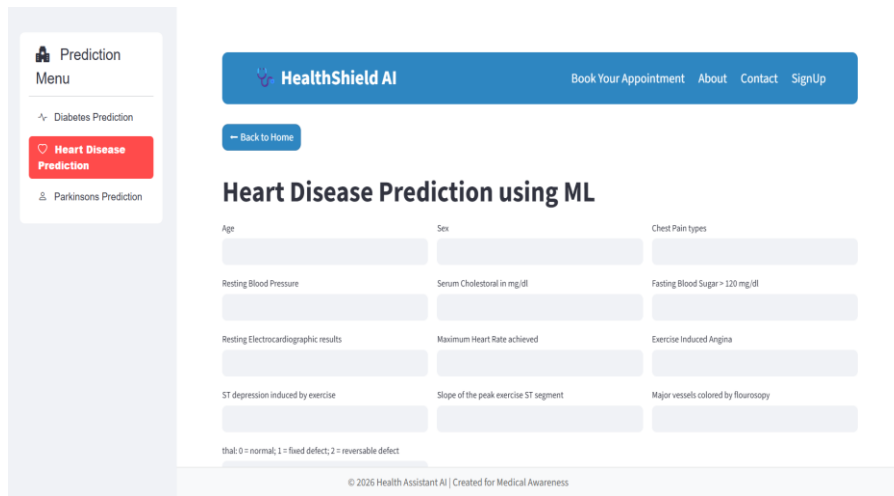


Figure 5: The Diabetes Prediction Home Page.

8.2 Heart Disease Prediction

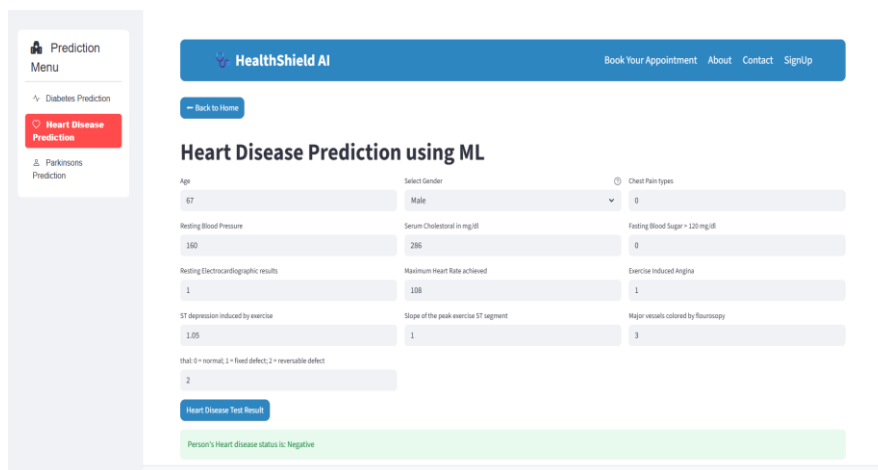


Figure 7: The Heart Disease Prediction Result Page

8.3 Parkinson's Prediction

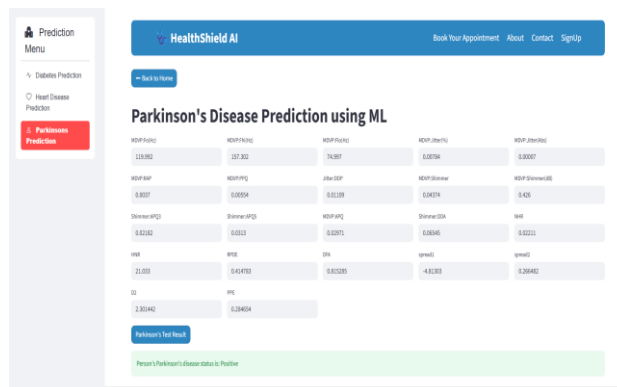


Figure 10: The Parkinson's Disease Prediction Result Page



8.4 Appointment Booking

After prediction, the system guides users to book an appointment with specialized doctors, enabling further consultation, proper diagnosis, and timely medical advice based on the generated health prediction results.

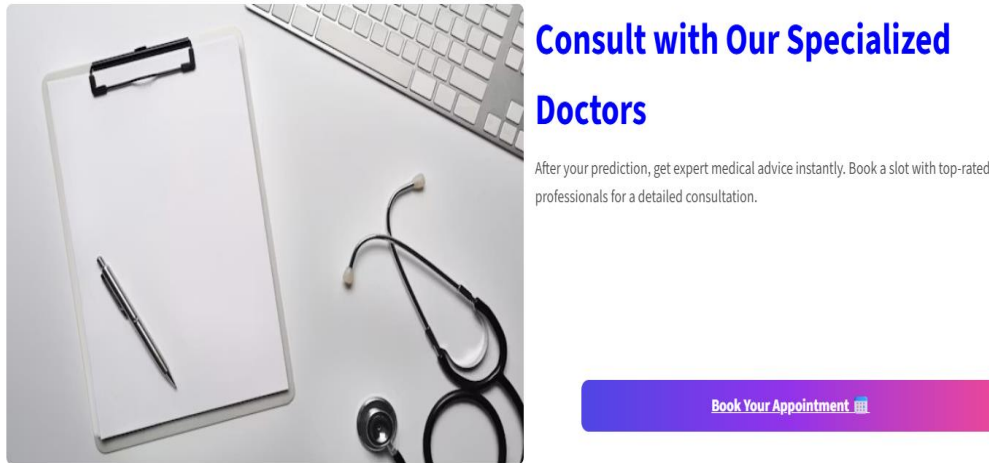


Figure 11: Post-Prediction Doctor Consultation and Appointment Booking Interface

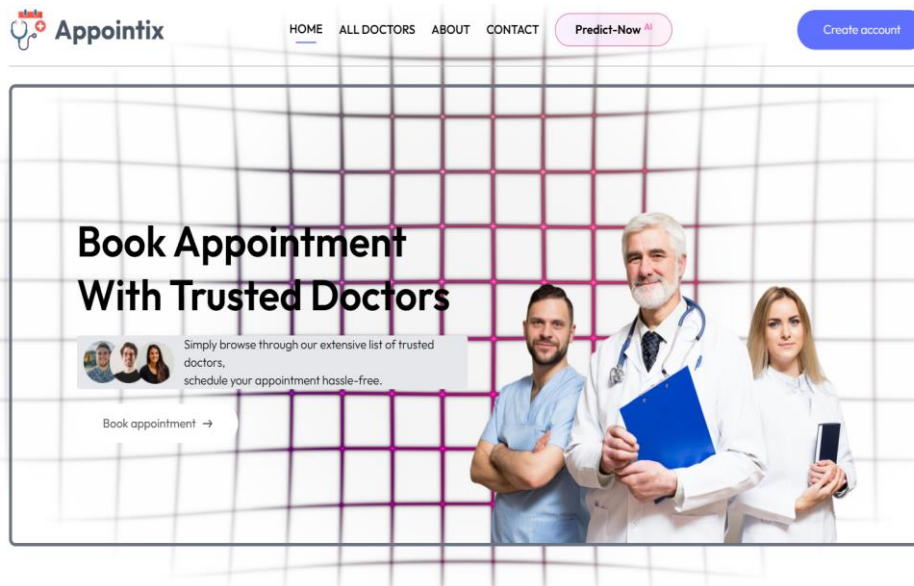


Figure 12: Doctor Appointment Booking Interface

CONCLUSION

In conclusion, the developed Multiple Disease Prediction System successfully demonstrates the practical application of machine learning in the healthcare domain. The system integrates three predictive modules—Diabetes, Heart Disease, and Parkinson's—into a unified and user-friendly Streamlit-based interface, enabling users to input clinical parameters and receive instant predictive outcomes. The implementation utilizes a combination of Linear Support Vector Machine and Logistic Regression models, ensuring flexibility and stability in prediction performance.



The project effectively highlights how machine learning models can assist in early disease detection by analyzing numerical health data. Although the exact training datasets and evaluation metrics are unspecified, the system follows a standard pipeline involving data input, preprocessing through type conversion, model inference, and result visualization.

The system follows a role-based workflow where users can register, log in, and either directly book appointments or first predict diseases using machine learning models. Admin manages doctors to ensure authenticity. This dual approach improves user convenience by supporting both early health prediction and immediate consultation, making the system efficient, flexible, and suitable for real-world healthcare applications.

REFERENCES

- [1]. C. T. Kor et al., "Explainable machine learning model for predicting first-time acute exacerbation in patients with chronic obstructive pulmonary disease", *Journal of Personalized Medicine*, vol. 12, no. 2, pp. 228, 2022
- [2]. Y. Perwej, Dr. Faiyaz Ahamad, Dr. Mohammad Zunnun Khan, N. Akhtar, "An Empirical Study on the Current State of Internet of Multimedia Things (IoMT)", *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, ISSN (Online) 2394 2320, Volume 8
- [3]. Littell, C.L., "Innovation in medical technology: Reading the indicators", *Health Affairs*, Vol. 13, No. 3, (1994), <https://doi.org/10.1377/hlthaff.13.3.226>
- [4]. Y. Perwej, Mohammed Y. Alzahrani, F. A. Mazarbhuiya, Md. Husamuddin, "The State-of-the-Art Cardiac Illness Prediction Using Novel Data Mining Technique", *International Journal of Engineering Sciences & Research Technology (IJESRT)*, ISSN: 2277-9655, Volume 7, Issue 2, Pages 725-739, 2018, DOI:
- [5]. Mobeen, A., Shafiq, M., Aziz, M.H. and Mohsin, M.J., "Impact of workflow interruptions on baseline activities of the doctors working in the emergency department", *BMJ Open Quality*, Vol. 11, No. 3, 2022
- [6]. Ahmed, S., Szabo, S. and Nilsen, K., "Catastrophic healthcare expenditure and impoverishment in tropical deltas: Evidence from the mekong delta region", *International Journal for Equity in Health*, Vol. 17, No. 1, 1 13, 2018
- [7]. Roberts, M.A. and Abery, B.H., "A person centered approach to home and community based services outcome measurement", *Frontiers in rehabilitation Sci.*, Vol. 4, 2023
- [8]. Y. Perwej, "An Evaluation of Deep Learning Miniature Concerning in Soft Computing", *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, ISSN (Online): 2278-1021, ISSN (Print): 2319-5940, Volume 4, Issue 2, Pages 10 - 16, 2015, DOI: 10.17148/IJARCCE.2015.4203
- [9]. M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease prediction by machine learning over big data from healthcare communities", *Ieee Access*, vol. 5, pp. 8869-8879, 2017
- [10]. M. A. E. Van Stiphout, J. Marinus, J. J. Van Hilten, F. Lobbezoo, and C. De Baat, "Oral health of Parkinson's disease patients: a case control study," *Parkinson's Disease*, vol. 2018, Article ID 9315285, 8 pages, 2018