Multi Disease Prediction Using Machine Learning

Buddhi Akhil, ECE ,Institute of Aeronautical Engineering, Hyderabad, India
22951A0412@iare.ac.in

Dr. S China Venkateshwarlu,Professor of ECE ,Institute of Aeronautical Engineering, Hyderabad, India
c.venkateshwarlu@iare.ac.in

Dr. V Siva Nagaraju,Professor of ECE ,Institute of Aeronautical Engineering, Hyderabad, India
v.sivanagaraju@iare.ac.in

Mr.U.Somanaidu [4],Asst. Professor of ECE ,Institute of Aeronautical Engineering, Hyderabad, India
u.somanaidu@iare.ac.in

Abstract:

Multiple Disease Prediction using Machine Learning, Deep Learning, and Streamlit is a comprehensive project aimed at predicting diseases such as **diabetes**, **heart disease**, and **Parkinson's disease**. The system leverages a combination of machine learning and deep learning algorithms, including **TensorFlow with** Keras**,** Support Vector Machine **(SVM)**, and **Logistic Regression**, to build accurate and reliable prediction models. These models are trained on publicly available datasets and are deployed using Streamlit **Cloud**, utilizing the Streamlit library to provide an intuitive and user-friendly interface.

The application allows users to choose from the supported disease categories—heart disease, diabetes, and Parkinson's disease. Upon selecting a disease, users are prompted to enter specific medical parameters relevant to that condition. After submitting the required information, the system processes the inputs through the trained model and delivers a real-time prediction, indicating the likelihood of the disease being present.

To ensure robust performance, the models undergo preprocessing steps including feature encoding and scaling, improving the training effectiveness and generalizability of the system. Model evaluation is performed using standard accuracy metrics, and the results show promising predictive capabilities for each condition.

This project addresses the growing need for accessible, technology-driven health assessment tools. By combining advanced machine learning techniques with an easy-to-use interface, it empowers users to monitor their health status and take preventive measures. Future enhancements could include integration with real-time data from wearable devices, increasing prediction accuracy and enabling continuous health monitoring.

1 INTRODUCTION :

In today's digitally connected and health-conscious world, chronic and life-threatening diseases such as diabetes, heart disease, and Parkinson's disease remain major global health concerns. Factors such as sedentary lifestyles, poor dietary habits, genetic predisposition, and increasing life expectancy have significantly contributed to the growing prevalence of these conditions. With healthcare systems placing greater emphasis on early detection and preventive care, the need for intelligent, accessible tools that can aid in the timely diagnosis of multiple diseases has become increasingly critical.

To address this growing demand, we present an innovative **Multi-Disease Prediction System** powered by machine learning and deep learning technologies. This application is designed to assist both individuals and healthcare professionals by providing accurate, early-stage predictions based on user-supplied health metrics. By integrating various disease-specific models into a single, unified platform, the system offers a scalable, cost-effective, and non-invasive solution for preliminary health assessments.

1.1 Understanding Disease-Specific Health Metrics:

At the core of this solution lies a collection of well-curated datasets for each target disease namely heart disease, diabetes, and Parkinson's disease. Each dataset comprises medically relevant attributes such as blood glucose levels, age, gender, blood pressure, cholesterol, motor function indicators, and more. These parameters serve as critical features for identifying disease-specific patterns. Using this data, the system applies machine learning algorithms such as **Support Vector Machines (SVM)**, **Logistic Regression**, and **TensorFlow-based deep learning models**—to deliver high-accuracy predictions tailored to each disease.

1.2 Real-Time Prediction and User Interaction:

Developed in Python and deployed through the **Streamlit** framework, the application features an intuitive, web-based interface. Users can select a disease from the available options and input relevant clinical data in real time. The system performs necessary preprocessing, such as feature encoding and scaling, and then applies the trained model to produce a prediction instantly indicating the presence or absence of the selected disease.

1.3 Insightful Feedback and Decision Support:

To ensure clarity and promote user engagement, the system presents its output using visual aids and clear textual explanations. These insights help users understand their current risk levels and support healthcare professionals in making informed, data-driven decisions. The visual feedback also encourages repeated use, allowing users to track changes in their health status over time.

1.4 Toward a Smarter Preventive Healthcare Ecosystem:

By integrating predictive analytics into a multi-disease framework, this system exemplifies the potential of artificial intelligence in transforming digital healthcare. It empowers users with timely, personalized health insights and bridges the gap between individual awareness and professional medical evaluation.

2 LITERATURE SURVEY

The increasing global burden of chronic and non-communicable diseases has prompted the development of intelligent tools capable of identifying multiple health conditions through early detection. Diseases such as diabetes, heart disease,, Parkinson's disease among the leading causes of morbidity and mortality. Traditional diagnostic methods, while effective, often involve high costs, extended processing times, and access to specialized facilities. Consequently, machine learning (ML) and deep learning (DL) have emerged as promising alternatives, offering efficient, cost-effective, and non-invasive solutions for disease prediction based on clinical data. This literature survey examines advancements in multi-disease prediction using ML and DL, focusing on data sources, algorithms, model performance, integration techniques, interpretability, and challenges.

Data Sources and Datasets

Multi-disease prediction systems typically rely on structured clinical datasets collected from hospitals, research institutions, and publicly available platforms such as **Kaggle** and **UCI Machine Learning Repository**. These datasets encompass a wide range of patient features including age, gender, blood pressure, glucose levels, cholesterol, and lifestyle factors. Each disease has its own set of diagnostic parameters—for instance, motor symptoms for Parkinson's disease. Some systems also incorporate unstructured data extracted using **Natural Language Processing (NLP)** from clinical notes.

Emerging technologies like **wearable sensors** and **IoT devices** have introduced new modalities for real-time health monitoring, allowing the inclusion of dynamic data such as heart rate, oxygen saturation, and movement patterns. However, one common challenge across most datasets is **class imbalance**, where healthy cases dominate disease-positive cases, leading to skewed prediction outcomes. To address this, techniques such as **SMOTE (Synthetic Minority Over-sampling Technique)** and random undersampling are commonly applied. Additionally, preprocessing techniques like **data imputation**, **feature normalization**, and **dimensionality reduction** are used to prepare data for robust model training.

Machine Learning Techniques

A wide array of ML and DL techniques has been explored in the domain of multi-disease prediction. These include both classical algorithms and more advanced neural networks, often customized to suit the unique characteristics of each disease dataset.

- **Logistic Regression and Decision Trees**: Favored for their simplicity and interpretability, these models serve as effective baselines. Logistic Regression is commonly used for binary classification tasks such as heart disease detection, while Decision Trees offer visual decision-making logic. Their standalone use, however, may limit accuracy.
- **Support Vector Machines (SVM)**: Known for high classification accuracy and the ability to handle non-linear relationships using kernel functions, SVMs are effective in diseases like Parkinson's and diabetes. However, they can become computationally expensive with large datasets.
- Random Forest and Ensemble Methods**:** Random Forests aggregate multiple decision trees to boost performance and reduce overfitting**.** Gradient Boosting techniques**, such as** XGBoost **and** LightGBM**,** further improve accuracy by correcting errors sequentially. Voting classifiers and stacking ensembles combine various models to achieve superior results across disease types.
- Deep Learning and Hybrid Models: For more complex diseases such as breast cancer and kidney disease, **Convolutional Neural Networks (CNNs)** and **Recurrent Neural Networks (RNNs)** have been employed. These models can learn high-level features from structured data or even medical imaging and time-series data. DL methods are powerful but require substantial training data and computational resources.

Feature Selection and Optimization

Effective feature selection is essential for model efficiency and performance. Techniques such as **mutual information gain**, **recursive feature elimination (RFE)**, and **L1/L2 regularization** help in reducing irrelevant or redundant features. Furthermore, **genetic algorithms** and **particle swarm optimization** have been applied for automated feature tuning.

Hyperparameter optimization plays a crucial role in fine-tuning models for optimal performance. Approaches like **grid search**, **random search**, **Bayesian optimization**, and **autoML tools** automate the process of discovering the best parameter configurations.

Model Interpretability and Explainability

Multi-disease prediction systems are often integrated into web or mobile platforms for easy accessibility. Frameworks such as **Streamlit**, **Flask**, and **Django** are commonly used to build interactive interfaces that allow users to input disease-specific parameters. Systems designed using **Streamlit Cloud** provide a cloud-hosted environment that simplifies deployment and user access.

Challenges and Limitations

Despite impressive advancements, multi-disease prediction systems face several challenges:

- **Limited and Homogeneous Datasets**: Most publicly available datasets are small, outdated, or lack demographic diversity, reducing the generalizability of models.
- Class imbalance can lead to high false negatives, particularly for rare diseases.
- Overfitting on small datasets is common, while bias can creep in if the training data is not representative.
- Deep models often offer better performance but lack interpretability.
- Bridging the gap between prediction tools and Electronic Health Records (EHR) or hospital workflows remains complex.
- Compliance with healthcare regulations like HIPAA or GDPR is necessary for real-world deployment.

These factors highlight the gap between research prototypes and real-world clinical applications.

Future Directions

Future advancements in multi-disease prediction will depend heavily on interdisciplinary collaboration among data scientists, healthcare professionals, and ethical experts. The integration of clinical knowledge is crucial for effective feature selection and rigorous model validation, ensuring that predictions align with real-world diagnostic needs. Federated learning presents a promising avenue for building robust, privacy-preserving models by enabling collaborative training across institutions without sharing sensitive patient data. The incorporation of multi-modal and multi-omics data—including genomic, proteomic, and lifestyle information—can enhance prediction accuracy and provide more personalized insights. Additionally, the integration of continuous health data from wearable devices and IoT sensors will support real-time monitoring and dynamic risk assessment, moving toward proactive healthcare systems. To gain clinical trust and ensure deployment readiness, future models must emphasize interpretability through explainable AI frameworks. Finally, ethical considerations such as algorithmic fairness, transparency, and regulatory compliance must be central to system design to promote equitable, responsible, and safe implementation in healthcare environments.

Table 2.1: literature survey

AUTHOR	ALGORITHM/TECHNIQUE	METHODOLOGY	REMARKS/PROBLEM	MERITS
1: Martín Abadi, Ashish Agarwal, et al. (2015)	-TensorFlow -Support Vector Machine (SVM) -Logistic Regression	Developed a large – scale machine learning system for heterogeneous environments,widely adopted for training deep learning models in healthcare	Random Requires significant computational resources for large models.	Scalable, supports deep learning architectures strong industry adoption
1:Corinna Cortes 2:Vladimir Vapnik June 2023	-Deep Neural Network (DNN) -Convolutional Neural Network (CNN) for ECG signal analysis	Used ECG signals and clinical data. Preprocessed signals and clinical features, trained DNN and CNN for classification of heart disease risk.	Linear decision boundary limits accuracy in complex datasets.	Simple, interpretable, widely used in healthcare analytics.

1.Zhang, Y., & Ghorbani, A. (2019) 2:Arora, P., Chaudhary, S., & Rana, M. (2020) 3. Saeed, A., & Al-Jumaily, A. (2020)	-ML for Heart Disease -ML for Diabetes - ML for Parkinson's	Used patient demographic and clinical data. Trained KNN and Naive Bayes classifiers to predict heart disease,Diabetes, Parkinson's Disease risk. Evaluated with accuracy, precision, recall metrics.	Limited dataset availability for handwriting data.	Novel input modality, promising accuracy with DL techniques.

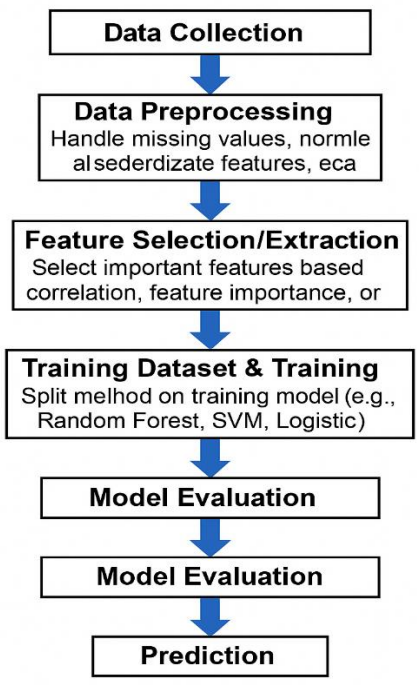

Figure 2.1: Existing block diagram for Multi Disease Prediction Using Machine Learning

2.1.1. **Data Collection**

This is the initial step where relevant patient data is collected for multiple diseases such as diabetes, heart disease and Parkinson's disease. The data may include medical history, clinical test results, lifestyle factors, demographic information, and physiological measurements.

2.1.2.**Data Preprocessing**

Preprocessing ensures the data is clean and consistent across all disease datasets. This involves handling missing or null values, removing duplicates, normalizing or standardizing numerical features, and encoding categorical variables to prepare the data for machine learning algorithms.

2.1.3.**Feature Selection/Extraction**

For each disease, significant features are selected using techniques such as correlation analysis, feature importance ranking, recursive feature elimination, or domain-specific medical knowledge. This step helps reduce noise and improve model performance by focusing on the most relevant attributes.

2.1.4.**Training Dataset & Training**

The processed data is split into training and testing sets. Various machine learning and deep learning models such as Support Vector Machines (SVM), Logistic Regression, Random Forest, and TensorFlow/Keras-based neural networks are trained separately for each disease to optimize prediction accuracy.

2.1.5.**Model Evaluation**

Each disease-specific model is evaluated using appropriate performance metrics such as accuracy, precision, recall, and F1 score. This step ensures that the models are reliable and capable of generalizing to unseen data.

2.1.6 **Prediction**

After evaluation, the trained models are used to make predictions on new/unseen data. The system allows users to select a disease and input relevant parameters, upon which the corresponding model predicts whether the individual is likely affected by that disease.

2.2 Problem statement

Despite significant advancements in machine learning and deep learning, the development of reliable and generalizable multi-disease prediction systems continues to face numerous challenges. While current models can predict individual diseases such as diabetes, heart disease, or Parkinson's with reasonable accuracy, integrating multiple disease classifiers into a unified framework remains a complex task due to differences in data characteristics, feature requirements, and diagnostic criteria across conditions One of the primary limitations lies in dataset variability. Publicly available datasets for diseases like diabetes, Parkinson's vary greatly in size, quality, and feature representation. These datasets often suffer from issues such as missing values, class imbalance, and inconsistent labeling standards. Additionally, many datasets are static and fail to capture longitudinal or temporal aspects of disease progression, limiting the predictive capacity for conditions that evolve over time.

Another challenge is the reliance on conventional machine learning models such as Logistic Regression, Support Vector Machines (SVM), and Decision Trees. While interpretable and computationally efficient, these models may not adequately capture the non-linear, high-dimensional interactions present in complex clinical data. Deep learning techniques, such as neural networks, offer greater representational power but introduce new concerns regarding

overfitting, computational overhead, and lack of interpretability—especially when applied to small or domain-specific datasets.

Furthermore, multi-disease prediction systems often face difficulties in handling heterogeneous input features across diseases. For example, Parkinson's prediction may rely on motor function indicators, while diabetes uses metabolic data models focus on blood chemistry. Designing a modular yet integrated system that can dynamically adjust to different input features without compromising accuracy remains a significant technical barrier.

Interpretability and clinical acceptance also present major hurdles. Models that function as black boxes—common with ensemble methods and deep learning—lack the transparency required for real-world clinical use. Clinicians need clear explanations of model decisions, especially when the stakes involve multiple diagnoses. Without adequate explainability, trust in AI-driven systems is diminished.

Interpretability and clinical acceptance also present major hurdles. Models that function as black boxes—common with ensemble methods and deep learning—lack the transparency required for real-world clinical use. Clinicians need clear explanations of model decisions, especially when the stakes involve multiple diagnoses. Without adequate explainability, trust in AI-driven systems is diminished.

Finally, the deployment of real-time, multi-disease systems is hampered by computational constraints and integration challenges with existing electronic health records (EHRs). These limitations underscore the need for scalable, explainable, and clinically viable AI models capable of providing accurate and accessible predictions for multiple diseases in a single platform.

proposed block diagram

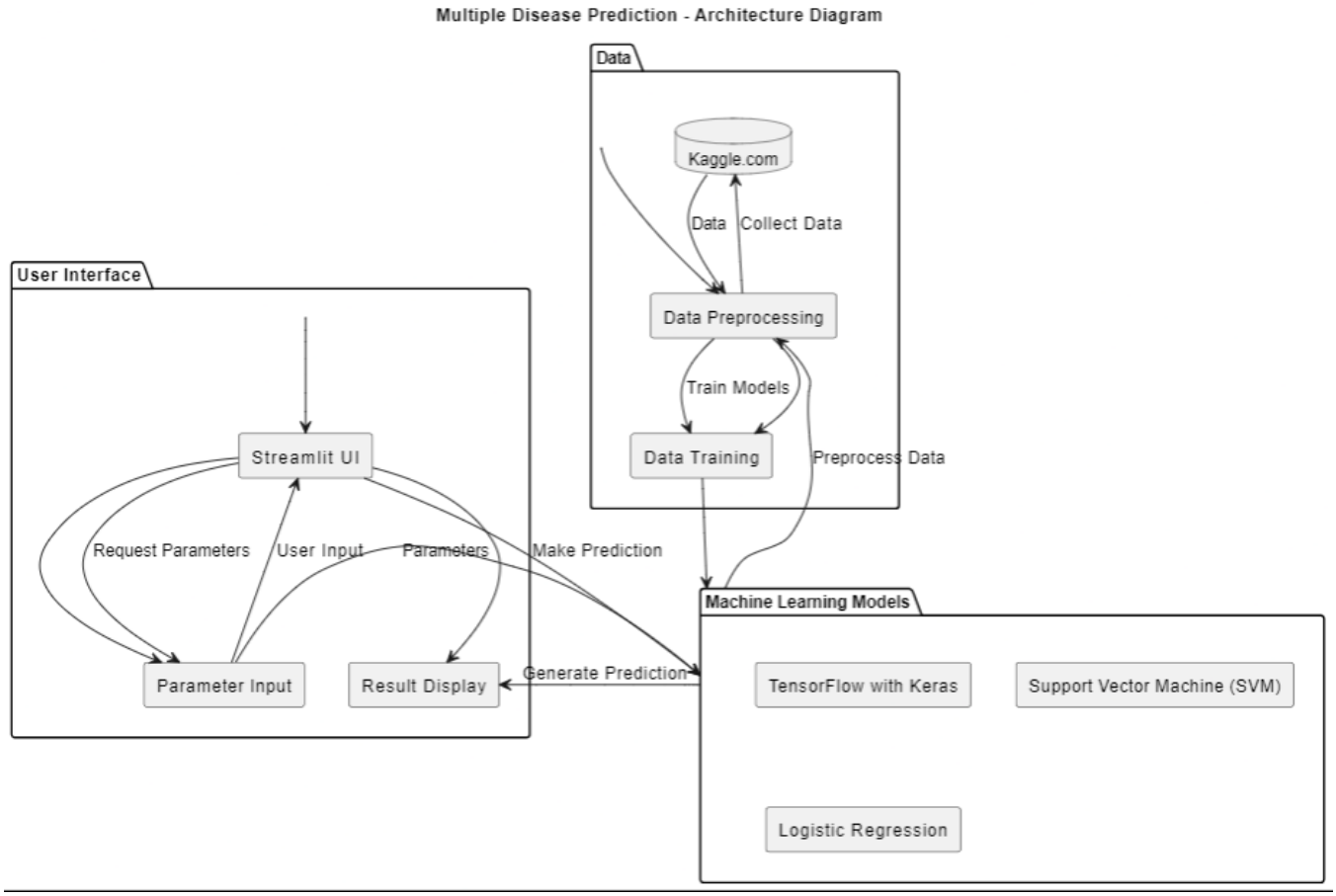

Figure 2.2: Heart disease prediction using Machine Learning

2.3 Software and Tools used :

2.3.1. Python

2.3.2. Visual Studio Code (VS Code)

Python is the primary programming language used in the Multiple Disease Prediction project due to its simplicity, readability, and rich ecosystem of data science and web development libraries. It supports the entire machine learning pipeline—from data preprocessing and model training to deployment and real-time prediction. In this project, key Python libraries such as **Streamlit**, **pickle**, and **streamlit-option-menu** are employed to create a robust, interactive, and user-friendly application.

Streamlit serves as the core framework for developing the web interface, enabling users to seamlessly select different disease prediction modules and input relevant medical parameters. The **streamlit-option-menu** library enhances the user experience by organizing the interface into a clean, navigable menu with disease-specific options such as diabetes,

heart disease and Parkinson's disease. Meanwhile, the **pickle** module is used to serialize and load the trained machine learning models efficiently, ensuring quick and accurate predictions without the need to retrain models during runtime.

Python's modular structure and integration capabilities significantly accelerate development and make the application accessible to both technical users and healthcare professionals. This powerful combination of tools and frameworks positions Python as the ideal language for implementing reliable, scalable, and easy-to-use health diagnostics solutions.

2.3.2 Visual Studio Code (VS Code)

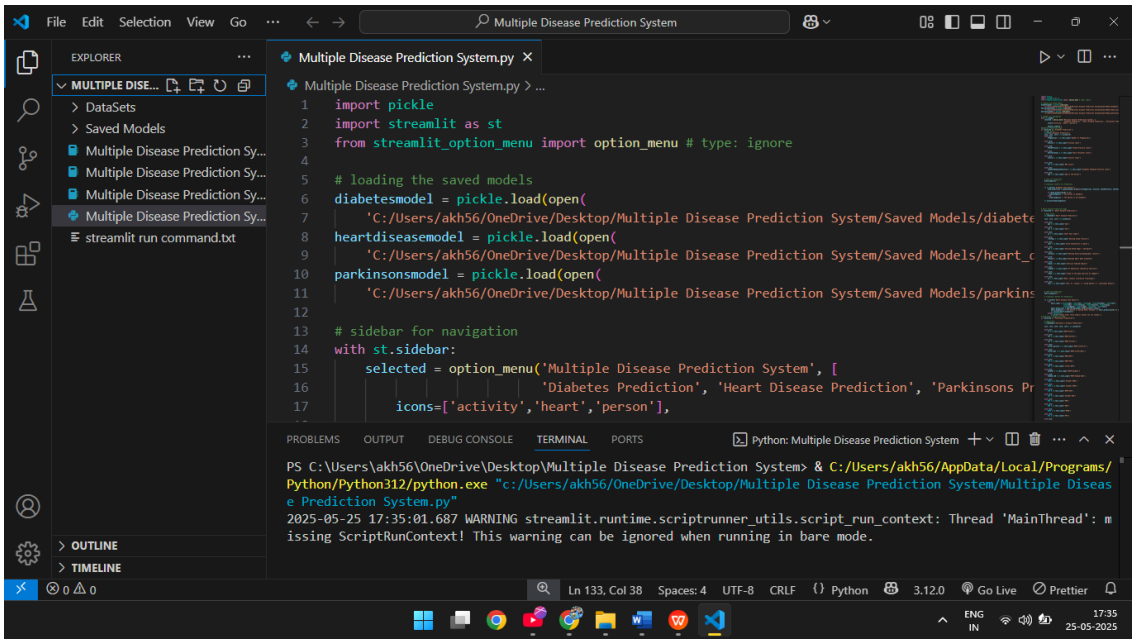

Figure 2.3: Visual Studio Code

In my **Multiple Disease Prediction** project code, you used the following **machine learning algorithms**:

1. Random Forest Classifier

Library: sklearn.ensemble.RandomForestClassifier

Description:

An **ensemble learning method** that builds multiple decision trees and merges their results (majority voting) for more accurate and stable predictions.

Why it's used:

- Handles both numerical and categorical features well
- Reduces overfitting
- Robust and accurate for classification tasks such as heart disease prediction

2. Logistic Regression

- **Library:** sklearn.linear_model.LogisticRegression
- **Description:**
  A **linear classification algorithm** used for binary classification problems (such as "Heart Disease" or "Not").
- **Why it's used:**
  - Simple and interpretable
  - Serves as a reliable baseline model
  - Performs well when there is a linear relationship between features and the target variable

Both models are trained and used for prediction, and I display the result from both on Streamlit app

2.3.3 Visual Studio Code

Visual Studio Code (VS Code) is an excellent code editor for machine learning projects, offering powerful tools and extensions that streamline the development workflow. It supports popular ML libraries like TensorFlow, PyTorch, and Pickle, and integrates seamlessly with Jupyter Notebooks for interactive coding. Features like IntelliSense, Git integration, and debugging make it easier to write, test, and manage ML code efficiently. With its flexibility and user-friendly interface, VS Code is a go-to choice for both beginners and professionals working on machine learning projects.

2.3.4. PowerShell / CMD

Used to run Python scripts, activate virtual environments, install dependencies using pip, and manage the runtime execution of the application (e.g., running app.py, activating venv, etc.).

2.3.5. Streamlit

Streamlit is an open-source Python library that turns data scripts into interactive web applications. It is used in this project to build a user-friendly interface where users can input medical details and receive real-time predictions. Streamlit simplifies frontend development, allowing developers to create responsive dashboards and web apps using pure Python without needing HTML, CSS, or JavaScript.

2.3.6. Logging

The logging module is used for tracking events that happen during the execution of the code. It is configured to filter out lower-level logs (like informational messages), showing only important warnings or errors. This helps in debugging and monitoring the app's behavior without cluttering the console.

2.4 Practical setup

Hardware Requirements

Laptop/PC with built-in

Internet: Only required for initial setup (for installing packages/models).

Input

Dataset Name: Heart Disease Dataset (Cleveland Heart Disease Dataset – UCI Repository)

Description

The Heart Disease dataset used in this project is a structured collection of patient health records commonly sourced from the UCI Machine Learning Repository. This version (heart.csv) contains 303 patient records and includes 13 clinical features that are predictive of the presence or absence of heart disease.

Each record represents a patient's medical profile, and the task is a binary classification problem — predicting whether or not a person is likely to develop heart disease,.

The target variable is labeled as:

- 1: Heart disease detected
- 0: No heart disease
- 1: The person is Diabetic
- 0: The person is not Diabetic
- 1: The person has Parkinson disease
- 0: The person doesn’t have Parkinson disease

Diabetes Disease Dataset Feature Description

Features	**Description**
Pregnancies	Number of Pregnancies
Glucose	Glucose Level
BP	Blood Pressure Level
Skin Thickness	Skin Thickness value

insulin	Insulin Level
BMI	BMI value
DPFV	Diabetes Pedigree Function Value
Age	Age of the Person

Heart Disease Dataset Feature Description

Feature	Description
age	Age of the patient in years
sex	Sex (1 = male; 0 = female)
cp	Chest pain type (0–3)
trestbps	Resting blood pressure (in mm Hg)
chol	Serum cholesterol in mg/dl
fbs	Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
restecg	Resting electrocardiographic results (0 = normal, 1 = ST-T abnormality, 2 = LV hypertrophy)
thalach	Maximum heart rate achieved
exang	Exercise-induced angina (1 = yes; 0 = no)

oldpeak	ST depression induced by exercise relative to rest
slope	Slope of the peak exercise ST segment (0 = up, 1 = flat, 2 = down)
ca	Number of major vessels (0–3) colored by fluoroscopy
thal	Thalassemia (3 = normal; 6 = fixed defect; 7 = reversible defect)
target	Output class label (1 = disease, 0 = no disease)

Parkinson's Disease Dataset Feature Description

Feature	Description
MDVP:Fo(Hz)	MDVP: Fundamental Frequency
MDVP:Fhi(Hz)	MDVP: Fundamental Frequency - High (Hertz)
MDVP:Flo(Hz)	MDVP: Fundamental Frequency - Low (Hertz)
MDVP:Jitter(%)	Multi-Dimensional Voice Program: Jitter Percentage
MDVP:Jitter(Abs)	Multi-Dimensional Voice Program: Absolute Jitter
MDVP:RAP	Relative Average Perturbation
MDVP:PPQ	Pitch Period Perturbation Quotient

Jitter:DDP	Difference of Differences of Periods
MDVP:Shimmer	Shimmer (Amplitude Perturbation) measured by the Multi-Dimensional Voice Program
MDVP:Shimmer(dB)	Shimmer (Amplitude Perturbation) measured by the Multi-Dimensional Voice Program in Decibels
Shimmer:APQ3	Amplitude Perturbation Quotient over 3 periods
Shimmer:APQ5	Amplitude Perturbation Quotient (5-point average)
MDVP:APQ	Mel Frequency Cepstral Coefficients (MFCC) Derived Voice Parameter: Amplitude Perturbation Quotient (APQ)
Shimmer:DDA	Shimmer: Drift Diffusion Analysis

3 Implentation

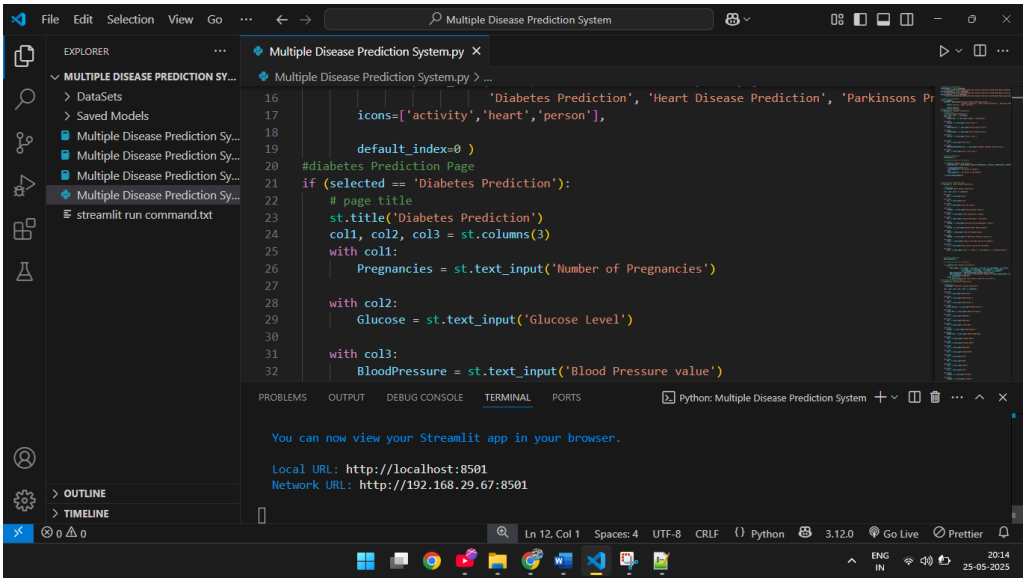

Figure 3.1: train model

Steps for implementation

1.Install Required VSoftware & Tools Install Required Software & Toolsss

2 Set Up a Virtual Environment

3 Install Dependencies(Streamlit pickle and etc)

4 Download & Preprocess the Dataset

4 Results and discussion

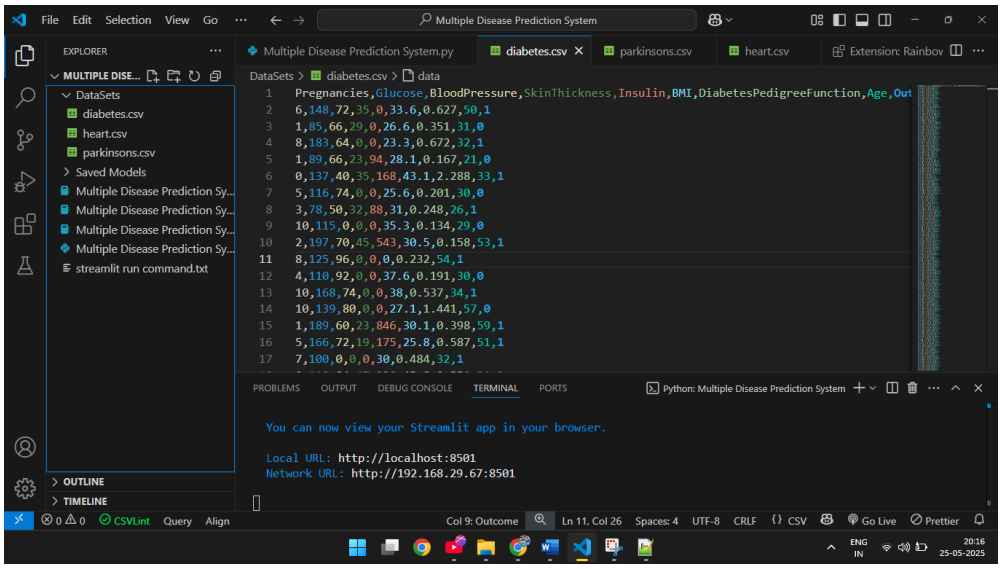

Figure 4.1: Dataset path

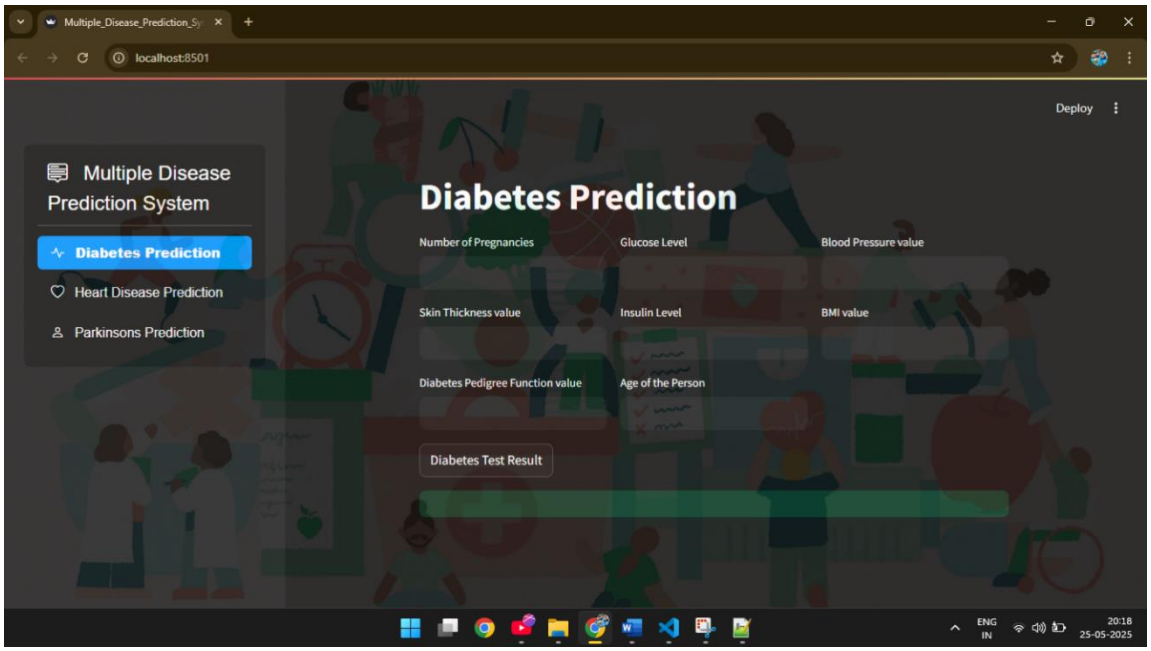

Figure 4.2: Model path

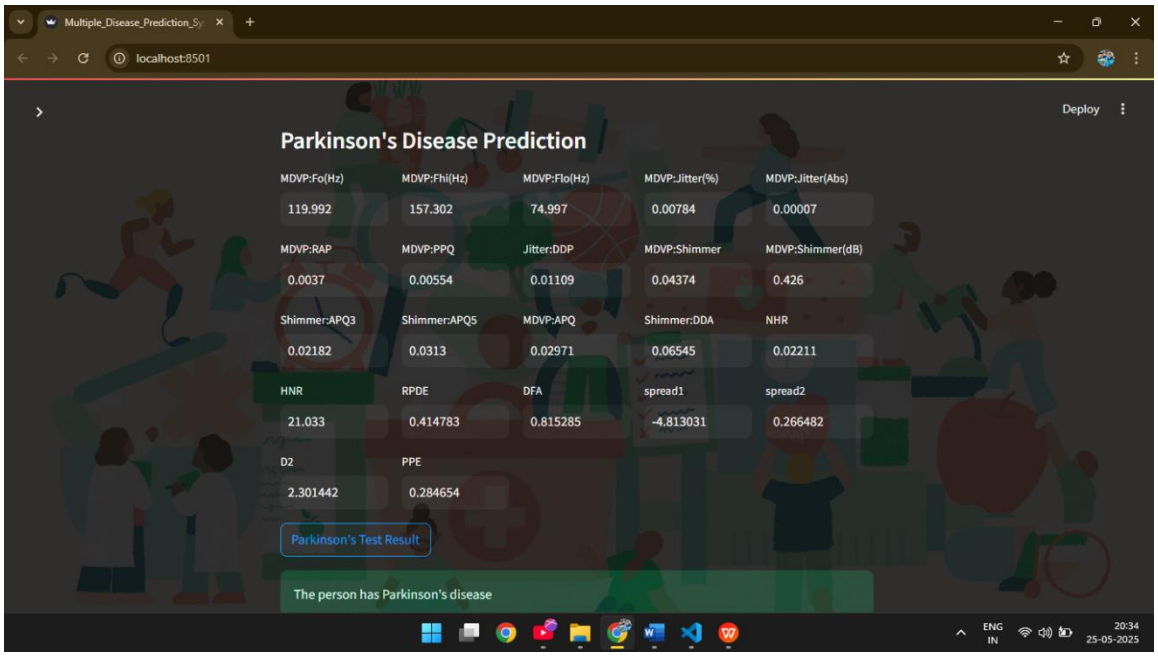

Figure 4.3: Output of Parkinson's disease prediction

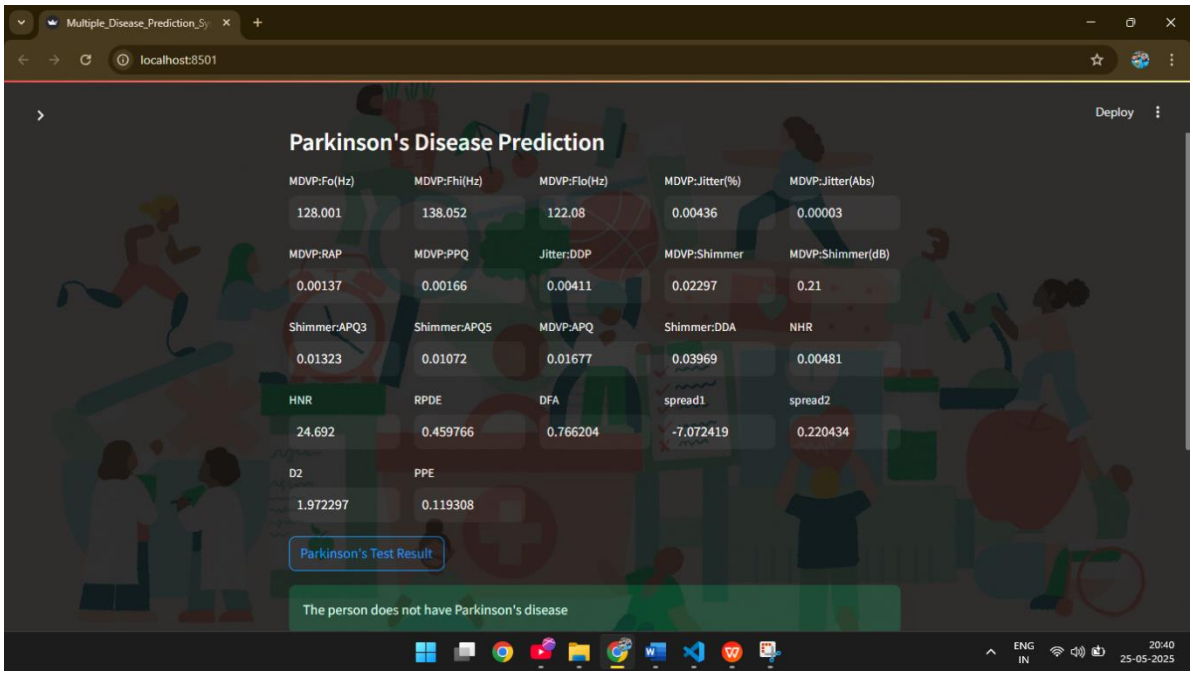

Figure 4.4: Output of No Parkinsons disease prediction

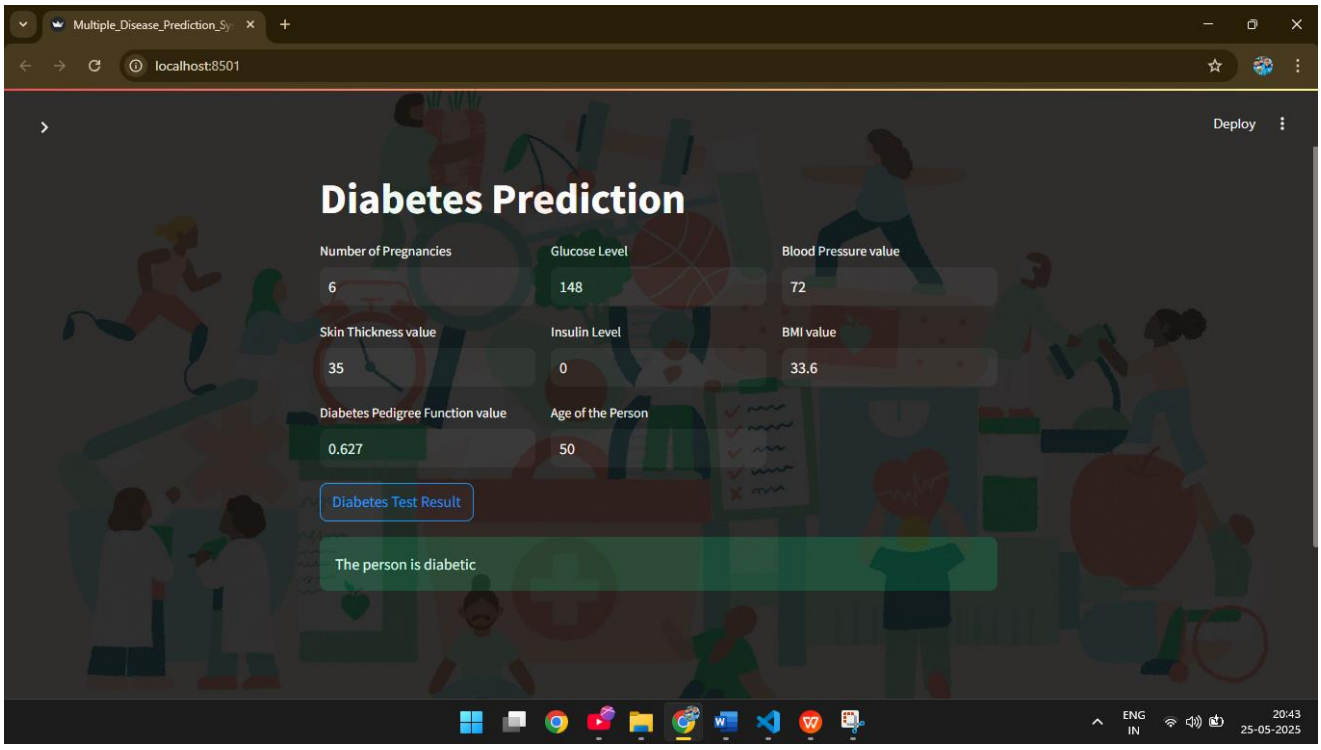

Figure 4.5: Output of diabetic prediction

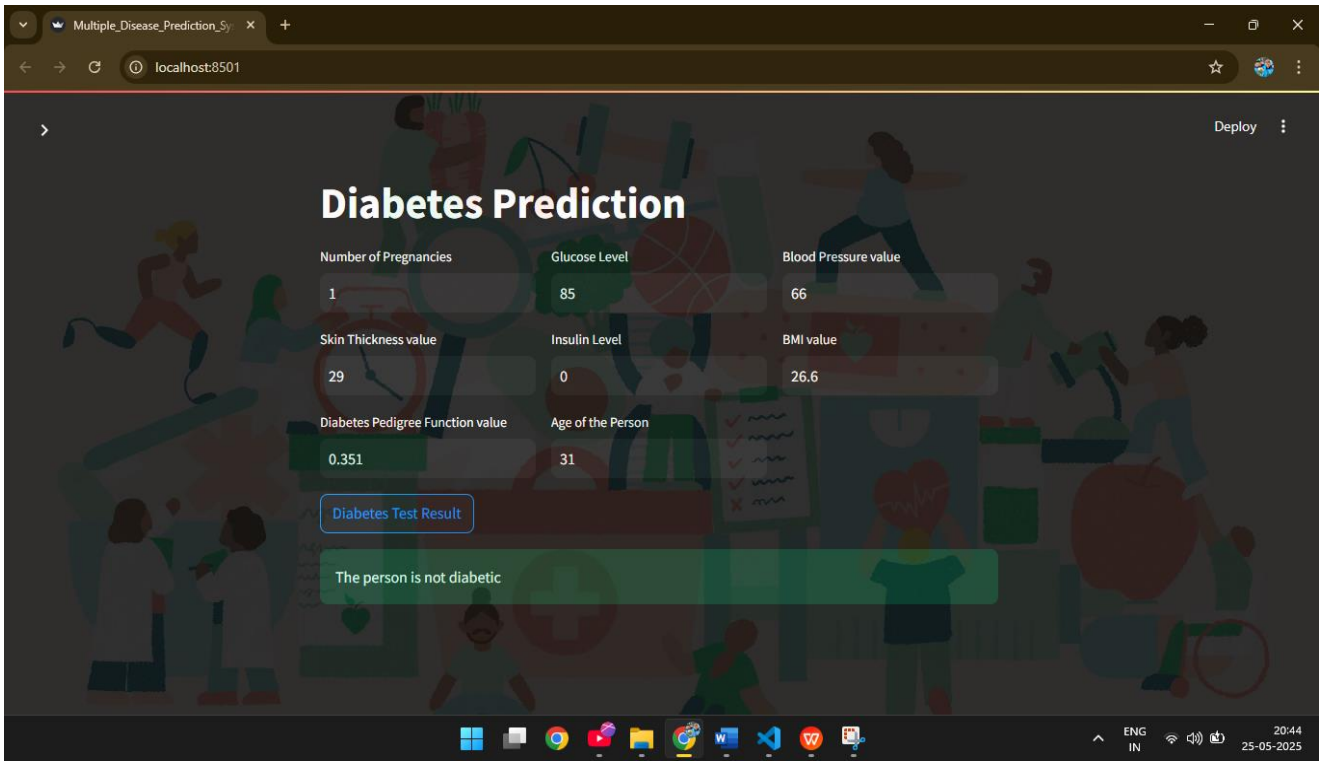

Figure 4.5: Output of Non Diabetic prediction

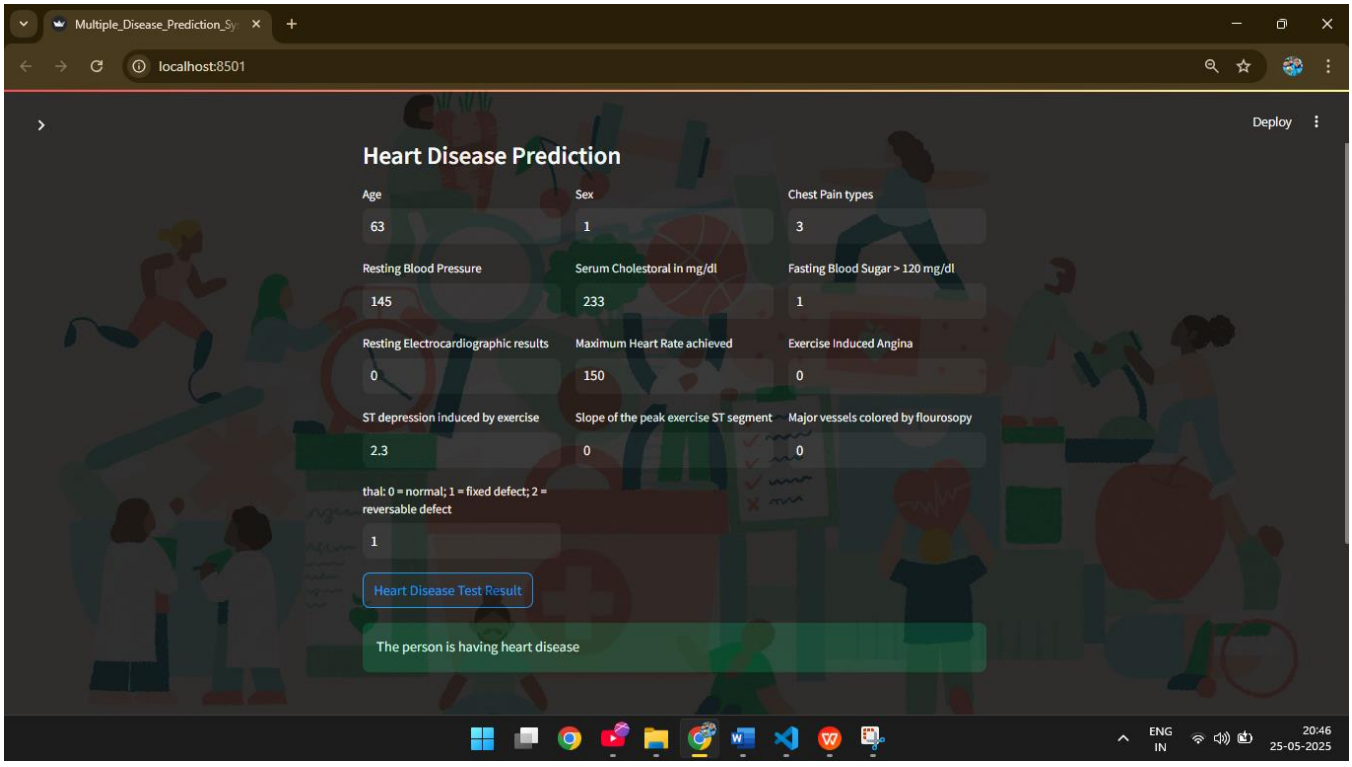

Figure 4.6: Output of Heart disease prediction

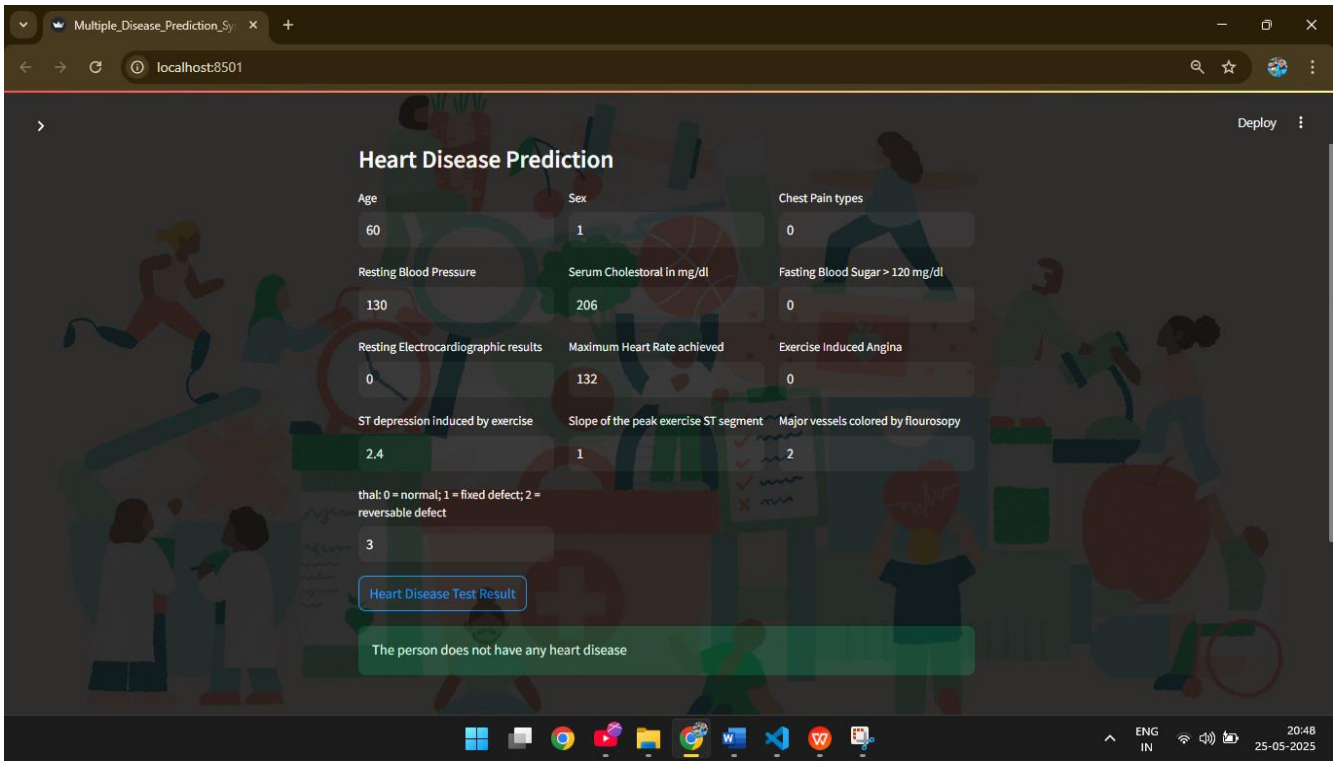

Figure 4.7: Output of Non Heart disease prediction

Table no 4.1;

Pregnancie	Glucose	BloodPress	SkinThickne	Insulin	BMI	DiabetesPe	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0
1	115	70	30	96	34.6	0.529	32	1
3	126	88	41	235	39.3	0.704	27	0

Table no 4.2;dgsfh

A	B	C	D	E	F	G	H	I	J	K	L	M	N
age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
58	0	3	150	283	1	0	162	0	1	2	0	2	1
50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
58	0	2	120	340	0	1	172	0	0	2	0	2	1
66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
59	1	0	135	234	0	1	161	0	0.5	1	0	3	1

Table no 4.3;dgsfh

name	MDVP:Fo(H	MDVP:Fhi(H	MDVP:Flo(H	MDVP:Jitte	MDVP:Jitte	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shin	MDVP:Shin	Shimmer:Al	Shimmer:Al	MDVP:APQ	Shimmer:D
phon_R01_	119.992	157.302	74.997	0.00784	0.00007	0.0037	0.00554	0.01109	0.04374	0.426	0.02182	0.0313	0.02971	0.06545
phon_R01_	122.4	148.65	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	0.626	0.03134	0.04518	0.04368	0.09403
phon_R01_	116.682	131.111	111.555	0.0105	0.00009	0.00544	0.00781	0.01633	0.05233	0.482	0.02757	0.03858	0.0359	0.0827
phon_R01_	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	0.517	0.02924	0.04005	0.03772	0.08771
phon_R01_	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	0.584	0.0349	0.04825	0.04465	0.1047
phon_R01_	120.552	131.162	113.787	0.00968	0.00008	0.00463	0.0075	0.01388	0.04701	0.456	0.02328	0.03526	0.03243	0.06985
phon_R01_	120.267	137.244	114.82	0.00333	0.00003	0.00155	0.00202	0.00466	0.01608	0.14	0.00779	0.00937	0.01351	0.02337
phon_R01_	107.332	113.84	104.315	0.0029	0.00003	0.00144	0.00182	0.00431	0.01567	0.134	0.00829	0.00946	0.01256	0.02487
phon_R01_	95.73	132.068	91.754	0.00551	0.00006	0.00293	0.00332	0.0088	0.02093	0.191	0.01073	0.01277	0.01717	0.03218
phon_R01_	95.056	120.103	91.226	0.00532	0.00006	0.00268	0.00332	0.00803	0.02838	0.255	0.01441	0.01725	0.02444	0.04324
phon_R01_	88.333	112.24	84.072	0.00505	0.00006	0.00254	0.0033	0.00763	0.02143	0.197	0.01079	0.01342	0.01892	0.03237
phon_R01_	91.904	115.871	86.292	0.0054	0.00006	0.00281	0.00336	0.00844	0.02752	0.249	0.01424	0.01641	0.02214	0.04272
phon_R01_	136.926	159.866	131.276	0.00293	0.00002	0.00118	0.00153	0.00355	0.01259	0.112	0.00656	0.00717	0.0114	0.01968
phon_R01_	139.173	179.139	76.556	0.0039	0.00003	0.00165	0.00208	0.00496	0.01642	0.154	0.00728	0.00932	0.01797	0.02184
phon_R01_	152.845	163.305	75.836	0.00294	0.00002	0.00121	0.00149	0.00364	0.01828	0.158	0.01064	0.00972	0.01246	0.03191
phon_R01_	142.167	217.455	83.159	0.00369	0.00003	0.00157	0.00203	0.00471	0.01503	0.126	0.00772	0.00888	0.01359	0.02316
phon_R01_	144.188	349.259	82.764	0.00544	0.00004	0.00211	0.00292	0.00632	0.02047	0.192	0.00969	0.012	0.02074	0.02908
phon_R01_	168.778	232.181	75.603	0.00718	0.00004	0.00284	0.00387	0.00853	0.03327	0.348	0.01441	0.01893	0.0343	0.04322
phon_R01_	153.046	175.829	68.623	0.00742	0.00005	0.00364	0.00432	0.01092	0.05517	0.542	0.02471	0.03572	0.05767	0.07413
phon_R01_	156.405	189.398	142.822	0.00768	0.00005	0.00372	0.00399	0.01116	0.03995	0.348	0.01721	0.02374	0.0431	0.05164
phon_R01_	153.848	165.738	65.782	0.0084	0.00005	0.00428	0.0045	0.01285	0.0381	0.328	0.01667	0.02383	0.04055	0.05
phon_R01_	153.88	172.86	78.128	0.0048	0.00003	0.00232	0.00267	0.00696	0.04137	0.37	0.02021	0.02591	0.04525	0.06062
phon_R01_	167.93	193.221	79.068	0.00442	0.00003	0.0022	0.00247	0.00661	0.04351	0.377	0.02228	0.0254	0.04246	0.06685
phon_R01_	173.917	192.735	86.18	0.00476	0.00003	0.00221	0.00258	0.00663	0.04192	0.364	0.02187	0.0247	0.03772	0.06562

Table no 4.4;dgsfh

name	MDVP:Fo(H	MDVP:Fhi(H	MDVP:Flo(H	MDVP:Jitte	MDVP:Jitte	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shin	MDVP:Shin	Shimmer:Al	Shimmer:Al	MDVP:APQ	Shimmer:D	NHR	HNR	status	RPDE	DFA
phon_R01_	119.992	157.302	74.997	0.00784	0.00007	0.0037	0.00554	0.01109	0.04374	0.426	0.02182	0.0313	0.02971	0.06545	0.02211	21.033	1	0.414783	0.815285
phon_R01_	122.4	148.65	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	0.626	0.03134	0.04518	0.04368	0.09403	0.01929	19.085	1	0.458359	0.819521
phon_R01_	116.682	131.111	111.555	0.0105	0.00009	0.00544	0.00781	0.01633	0.05233	0.482	0.02757	0.03858	0.0359	0.0827	0.01309	20.651	1	0.429895	0.825288
phon_R01_	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	0.517	0.02924	0.04005	0.03772	0.08771	0.01353	20.644	1	0.434969	0.819235
phon_R01_	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	0.584	0.0349	0.04825	0.04465	0.1047	0.01767	19.649	1	0.417356	0.823484
phon_R01_	120.552	131.162	113.787	0.00968	0.00008	0.00463	0.0075	0.01388	0.04701	0.456	0.02328	0.03526	0.03243	0.06985	0.01222	21.378	1	0.415564	0.825069
phon_R01_	120.267	137.244	114.82	0.00333	0.00003	0.00155	0.00202	0.00466	0.01608	0.14	0.00779	0.00937	0.01351	0.02337	0.00607	24.886	1	0.59604	0.764112
phon_R01_	107.332	113.84	104.315	0.0029	0.00003	0.00144	0.00182	0.00431	0.01567	0.134	0.00829	0.00946	0.01256	0.02487	0.00344	26.892	1	0.63742	0.763262
phon_R01_	95.73	132.068	91.754	0.00551	0.00006	0.00293	0.00332	0.0088	0.02093	0.191	0.01073	0.01277	0.01717	0.03218	0.0107	21.812	1	0.615551	0.773587
phon_R01_	95.056	120.103	91.226	0.00532	0.00006	0.00268	0.00332	0.00803	0.02838	0.255	0.01441	0.01725	0.02444	0.04324	0.01022	21.862	1	0.547037	0.798463
phon_R01_	88.333	112.24	84.072	0.00505	0.00006	0.00254	0.0033	0.00763	0.02143	0.197	0.01079	0.01342	0.01892	0.03237	0.01166	21.118	1	0.611137	0.776156
phon_R01_	91.904	115.871	86.292	0.0054	0.00006	0.00281	0.00336	0.00844	0.02752	0.249	0.01424	0.01641	0.02214	0.04272	0.01141	21.414	1	0.58339	0.79252
phon_R01_	136.926	159.866	131.276	0.00293	0.00002	0.00118	0.00153	0.00355	0.01259	0.112	0.00656	0.00717	0.0114	0.01968	0.00581	25.703	1	0.4606	0.646846
phon_R01_	139.173	179.139	76.556	0.0039	0.00003	0.00165	0.00208	0.00496	0.01642	0.154	0.00728	0.00932	0.01797	0.02184	0.01041	24.889	1	0.430166	0.665833
phon_R01_	152.845	163.305	75.836	0.00294	0.00002	0.00121	0.00149	0.00364	0.01828	0.158	0.01064	0.00972	0.01246	0.03191	0.00609	24.922	1	0.474791	0.654027
phon_R01_	142.167	217.455	83.159	0.00369	0.00003	0.00157	0.00203	0.00471	0.01503	0.126	0.00772	0.00888	0.01359	0.02316	0.00839	25.175	1	0.565924	0.658245
phon_R01_	144.188	349.259	82.764	0.00544	0.00004	0.00211	0.00292	0.00632	0.02047	0.192	0.00969	0.012	0.02074	0.02908	0.01859	22.333	1	0.56738	0.644692
phon_R01_	168.778	232.181	75.603	0.00718	0.00004	0.00284	0.00387	0.00853	0.03327	0.348	0.01441	0.01893	0.0343	0.04322	0.02919	20.376	1	0.631099	0.605417
phon_R01_	153.046	175.829	68.623	0.00742	0.00005	0.00364	0.00432	0.01092	0.05517	0.542	0.02471	0.03572	0.05767	0.07413	0.0316	17.28	1	0.665318	0.719467
phon_R01_	156.405	189.398	142.822	0.00768	0.00005	0.00372	0.00399	0.01116	0.03995	0.348	0.01721	0.02374	0.0431	0.05164	0.03365	17.153	1	0.649554	0.68608
phon_R01_	153.848	165.738	65.782	0.0084	0.00005	0.00428	0.0045	0.01285	0.0381	0.328	0.01667	0.02383	0.04055	0.05	0.03871	17.536	1	0.660125	0.704087
phon_R01_	153.88	172.86	78.128	0.0048	0.00003	0.00232	0.00267	0.00696	0.04137	0.37	0.02021	0.02591	0.04525	0.06062	0.01849	19.493	1	0.629017	0.698951
phon_R01_	167.93	193.221	79.068	0.00442	0.00003	0.0022	0.00247	0.00661	0.04351	0.377	0.02228	0.0254	0.04246	0.06685	0.0128	22.468	1	0.61906	0.679834
phon_R01_	173.917	192.735	86.18	0.00476	0.00003	0.00221	0.00258	0.00663	0.04192	0.364	0.02187	0.0247	0.03772	0.06562	0.0184	20.422	1	0.537264	0.686894

Table no 4.5;dgsfh

MDVP:Jitte	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shin	MDVP:Shin	Shimmer:Al	Shimmer:Al	MDVP:APQ	Shimmer:D	NHR	HNR	status	RPDE	DFA	spread1	spread2	D2	PPE
0.00007	0.0037	0.00554	0.01109	0.04374	0.426	0.02182	0.0313	0.02971	0.06545	0.02211	21.033	1	0.414783	0.815285	-4.813031	0.266482	2.301442	0.284654
0.00008	0.00465	0.00696	0.01394	0.06134	0.626	0.03134	0.04518	0.04368	0.09403	0.01929	19.085	1	0.458359	0.819521	-4.075192	0.33559	2.486855	0.368674
0.00009	0.00544	0.00781	0.01633	0.05233	0.482	0.02757	0.03858	0.0359	0.0827	0.01309	20.651	1	0.429895	0.825288	-4.443179	0.311173	2.342259	0.332634
0.00009	0.00502	0.00698	0.01505	0.05492	0.517	0.02924	0.04005	0.03772	0.08771	0.01353	20.644	1	0.434969	0.819235	-4.117501	0.334147	2.405554	0.368975
0.00011	0.00655	0.00908	0.01966	0.06425	0.584	0.0349	0.04825	0.04465	0.1047	0.01767	19.649	1	0.417356	0.823484	-3.747787	0.234513	2.33218	0.410335
0.00008	0.00463	0.0075	0.01388	0.04701	0.456	0.02328	0.03526	0.03243	0.06985	0.01222	21.378	1	0.415564	0.825069	-4.242867	0.299111	2.18756	0.357775
0.00003	0.00155	0.00202	0.00466	0.01608	0.14	0.00779	0.00937	0.01351	0.02337	0.00607	24.886	1	0.59604	0.764112	-5.634322	0.257682	1.854785	0.211756
0.00003	0.00144	0.00182	0.00431	0.01567	0.134	0.00829	0.00946	0.01256	0.02487	0.00344	26.892	1	0.63742	0.763262	-6.167603	0.183721	2.064693	0.163755
0.00006	0.00293	0.00332	0.0088	0.02093	0.191	0.01073	0.01277	0.01717	0.03218	0.0107	21.812	1	0.615551	0.773587	-5.498678	0.327769	2.322511	0.231571
0.00006	0.00268	0.00332	0.00803	0.02838	0.255	0.01441	0.01725	0.02444	0.04324	0.01022	21.862	1	0.547037	0.798463	-5.011879	0.325996	2.432792	0.271362
0.00006	0.00254	0.0033	0.00763	0.02143	0.197	0.01079	0.01342	0.01892	0.03237	0.01166	21.118	1	0.611137	0.776156	-5.24977	0.391002	2.407313	0.24974
0.00006	0.00281	0.00336	0.00844	0.02752	0.249	0.01424	0.01641	0.02214	0.04272	0.01141	21.414	1	0.58339	0.79252	-4.960234	0.363566	2.642476	0.275931
0.00002	0.00118	0.00153	0.00355	0.01259	0.112	0.00656	0.00717	0.0114	0.01968	0.00581	25.703	1	0.4606	0.646846	-6.547148	0.152813	2.041277	0.138512
0.00003	0.00165	0.00208	0.00496	0.01642	0.154	0.00728	0.00932	0.01797	0.02184	0.01041	24.889	1	0.430166	0.665833	-5.660217	0.254989	2.519422	0.199889
0.00002	0.00121	0.00149	0.00364	0.01828	0.158	0.01064	0.00972	0.01246	0.03191	0.00609	24.922	1	0.474791	0.654027	-6.105098	0.203653	2.125618	0.1701
0.00003	0.00157	0.00203	0.00471	0.01503	0.126	0.00772	0.00888	0.01359	0.02316	0.00839	25.175	1	0.565924	0.658245	-5.340115	0.210185	2.205546	0.234589
0.00004	0.00211	0.00292	0.00632	0.02047	0.192	0.00969	0.012	0.02074	0.02908	0.01859	22.333	1	0.56738	0.644692	-5.44004	0.239764	2.264501	0.218164
0.00004	0.00284	0.00387	0.00853	0.03327	0.348	0.01441	0.01893	0.0343	0.04322	0.02919	20.376	1	0.631099	0.605417	-2.93107	0.434326	3.007463	0.430788
0.00005	0.00364	0.00432	0.01092	0.05517	0.542	0.02471	0.03572	0.05767	0.07413	0.0316	17.28	1	0.665318	0.719467	-3.949079	0.35787	3.10901	0.377429
0.00005	0.00372	0.00399	0.01116	0.03995	0.348	0.01721	0.02374	0.0431	0.05164	0.03365	17.153	1	0.649554	0.68608	-4.554466	0.340176	2.856676	0.322111
0.00005	0.00428	0.0045	0.01285	0.0381	0.328	0.01667	0.02383	0.04055	0.05	0.03871	17.536	1	0.660125	0.704087	-4.095442	0.262564	2.73971	0.365391
0.00003	0.00232	0.00267	0.00696	0.04137	0.37	0.02021	0.02591	0.04525	0.06062	0.01849	19.493	1	0.629017	0.698951	-5.18696	0.237622	2.557536	0.259765
0.00003	0.0022	0.00247	0.00661	0.04351	0.377	0.02228	0.0254	0.04246	0.06685	0.0128	22.468	1	0.61906	0.679834	-4.330956	0.262384	2.916777	0.285695
0.00003	0.00221	0.00258	0.00663	0.04192	0.364	0.02187	0.0247	0.03772	0.06562	0.0184	20.422	1	0.537264	0.686894	-5.248776	0.210279	2.547508	0.253556

5 .Conclusion

The Multi Disease Prediction System using Machine Learning and Deep Learning demonstrates the potential of AI in early diagnosis of chronic diseases like diabetes, heart disease, and Parkinson's disease. By integrating diverse prediction models into a single platform and deploying them via a user-friendly web interface (Streamlit), the system offers a scalable and accessible solution for preliminary health assessments. The use of models like Random Forest, Logistic Regression, and TensorFlow ensures both reliability and flexibility in predictions. Overall, the project bridges the gap between complex medical data and user-friendly digital health tools.

6. Future Scope

- Integration with Wearable Devices: Real-time data from smartwatches and sensors can
  - enhance prediction accuracy and support continuous health monitoring.
  - Federated Learning: Allows collaboration across healthcare institutions without sharing
  - sensitive patient data, improving both privacy and model robustness.
  - Incorporation of Multi-Modal Data: Adding genomic, proteomic, and lifestyle data can
  - enable more personalized and comprehensive health insights.
  - Explainable AI (XAI): Enhancing model transparency will increase clinical trust and
  - support real-world adoption in hospitals and clinics.
  - Regulatory Compliance: Ensuring the system adheres to standards like HIPAA or
  - GDPR for legal and ethical deployment in healthcare environments.

References

1. TensorFlow: Martín Abadi, Ashish Agarwal, et al. (2015). TensorFlow: Large-scale machine learning on
heterogeneous systems. arXiv preprint arXiv:1603.04467.
2. Keras: François Chollet et al. (2015). Keras. GitHub repository.
3 .Support Vector Machine (SVM): Corinna Cortes and Vladimir Vapnik (1995). Support-vector networks. Machine Learning, 20(3), 273-297.
4.Logistic Regression: Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). John Wiley & Sons.
5. Streamlit: Streamlit Documentation. https://docs.streamlit.io/
6. Kaggle: Kaggle website. https://www.kaggle.com/
7. Data sources: You can provide the specific datasets you used from Kaggle.com, mentioning the

authors or contributors of the datasets.

8 .Zhang, Y., & Ghorbani, A. (2019). A review on machine learning algorithms for diagnosis of heart disease. IEEE Access, 7, 112751-112760.

9 .Arora, P., Chaudhary, S., & Rana, M. (2020). Prediction of diabetes using machine learning algorithms: A review. Journal of Ambient Intelligence and Humanized Computing, 11(6), 2575-2589.

10 .Kaur, H., Batra, N., & Rani, R. (2020). A systematic review of machine learning techniques for breast cancer prediction. Journal of Medical Systems, 44(11), 1-15.

11. Gupta, D., & Rathore, S. (2021). A comprehensive review on machine learning algorithms for kidney disease diagnosis. Journal of Medical Systems, 45(1), 1-17.

12. Saeed, A., & Al-Jumaily, A. (2020). Machine learning techniques for Parkinson's disease diagnosis using handwriting: A review. Computers in Biology and Medicine, 122, 103804.

Author credentials

Buddhi Akhil studying 3rd year department of Electronics And Communication Engineering at Institute Of Aeronautical Engineering ,Dundigal .. He Published a Research Paper Recently At IJSREM as a part of academics He has interest in VLSI.

Dr Sonagiri China Venkateswarlu professor in the Department of Electronics and Communication Engineering at the Institute of Aeronautical Engineering (IARE). He holds a Ph.D. degree in Electronics and Communication Engineering with a specialization in Digital Speech Processing. He has more than 40 citations and paper publications across various publishing platforms, and expertise in teaching subjects such as microprocessors and microcontrollers,digital signal processing, digital image processing, and speech processing. With 20 years of teaching experience, he can be contacted at email: **c.venkateswarlu@iare.ac.in**

Dr. V. Siva Nagaraju is a professor in the Department of Electronics and Communication Engineering at the Institute of Aeronautical Engineering (IARE). He holds a Ph.D. degree in Electronics and Communication Engineering with a specialization in Microwave Engineering. With over 21 years of academic experience, Dr. Nagaraju is known for his expertise in teaching core electronics subjects and has contributed significantly to the academic and research community. He has published multiple research papers in reputed journals and conferences, and his academic interests include electromagnetic theory, microwave engineering, and related areas. He can be contacted at email: **v.sivanagaraju@iare.ac.in**.

Mr.U.Somanaidu is an Assistant Professor in the Department of Electronics and Communication Engineering at the Institute of Aeronautical Engineering (IARE). He holds an M.Tech degree in VLSI and Embeded Systems from Jawaharlal Nehru Technological University, Hyderabad, and is currently pursuing his Ph.D. at Veltech (Deemed to be University), Chennai, Tamil Nadu. With nearly 14 Yearsof teaching experience, he specializes in VLSI and Embeded Systems and is actively involved in academic and research activities.
He has contributed to the academic community through his teaching and continues to enhance his knowledge and skills through ongoing doctoral research. He can be contacted at email: u.somanaidu@iare.ac.in