Heart Disease Prediction Using Machine Learning

Tammali Adithya Vardhan[1], ECE ,Institute of Aeronautical Engineering, Hyderabad, India 22951A0408@iare.ac.in

Dr. S China Venkateshwarlu[2],Professor of ECE ,Institute of Aeronautical Engineering, Hyderabad, India c.venkateshwarlu@iare.ac.in

Dr. V Siva Nagaraju[3],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:

Heart disease remains one of the most critical health concerns globally, often leading to life-threatening complications if not identified in its early stages. This project introduces a data-driven solution for predicting the likelihood of heart disease using machine learning techniques. By analyzing structured health data including patient attributes such as age, sex, blood pressure, cholesterol levels, chest pain type, and other clinical features the system learns patterns associated with the onset of heart-related conditions.

The predictive framework is developed using supervised learning algorithms, specifically Random Forest and Logistic Regression, which are trained on a publicly available heart disease dataset. Prior to model training, the data undergo preprocessing procedures such as feature scaling and encoding to enhance the learning process. The application is built using Python and incorporates libraries like scikit-learn for modeling, pandas for data manipulation, and Streamlit for the user-facing interface. This setup enables users to input health parameters via a graphical interface and receive real-time predictions regarding potential heart disease risk.

Model performance is evaluated using accuracy metrics and validated through testing on unseen data. The results indicate that the trained models can effectively identify individuals at risk, offering a practical approach for early detection. This system has potential applications in clinical decision support and personal health monitoring. Future developments may involve incorporating more diverse datasets and expanding the model to include real-time physiological data for improved accuracy.

1 INTRODUCTION :

In today's fast-paced and health-conscious world, cardiovascular diseases have emerged as one of the leading causes of death and disability across the globe. Factors such as unhealthy lifestyle habits, genetic predisposition, and environmental influences have significantly contributed to the rise in heart-related illnesses. As healthcare systems increasingly shift towards preventive care, the demand for intelligent tools that can detect and predict such conditions in their early stages has never been more vital.

To address this need, we introduce an innovative Heart Disease Prediction System powered by machine learning. This application is designed to support individuals and healthcare professionals by providing early warnings based on user-provided medical data. The system employs advanced data-driven algorithms to evaluate the risk of heart disease, offering a scalable, low-cost, and non-invasive method for initial screening.

1.1 Understanding Patient Health Metrics:

At the core of our solution lies a well-structured clinical dataset that includes key features such as age, gender, resting blood pressure, cholesterol levels, fasting blood sugar, chest pain type, and more. These features serve as vital indicators of heart health. Using this data, the application utilizes machine learning models—specifically Random Forest and Logistic Regression to identify complex patterns and make predictions with a high degree of accuracy.

1.2 Real-time Risk Analysis and Interpretation:

The application is developed using Python and integrated with **Streamlit**, offering an interactive, web-based interface. Users can input their personal health data in real time and receive immediate results indicating their potential risk for heart disease. The system processes the inputs, performs data standardization, and applies the trained model to generate a prediction.

1.3 Visual Feedback and Decision Support:

To ensure interpretability and enhance user engagement, the system presents results through graphical outputs and clear textual insights. These visualizations not only assist users in understanding their risk level but also help them monitor changes over time when used periodically. Healthcare providers can also leverage these insights as a preliminary assessment tool before more advanced diagnostics.

1.4 Towards Preventive Healthcare:

Our Heart Disease Prediction System aims to bridge the gap between patient awareness and clinical diagnostics. It empowers users to take proactive steps toward heart health management by offering personalized and timely information. Whether for personal health monitoring or clinical triage, this application exemplifies how machine learning can drive innovation in digital healthcare.

2 LITERATURE SURVEY

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, accounting for a significant proportion of deaths annually. Among these, heart disease, including coronary heart disease and heart failure, constitutes the majority of cases. Early and accurate prediction of heart conditions enables timely clinical intervention, improving patient outcomes. Traditional diagnostic techniques, such as electrocardiograms (ECG) and angiography, though effective, require specialized equipment, skilled personnel, and often involve time-consuming procedures. Machine learning (ML) provides a promising data-driven alternative by automatically identifying complex patterns in large clinical datasets that may be difficult for humans to detect. Recent advancements demonstrate that ML methods—including random forests, support vector machines, neural networks, and ensemble approaches—can achieve high accuracy in heart disease classification, encouraging extensive research into diverse ML methodologies. This survey reviews the literature on ML-based heart disease prediction, focusing on data sources, algorithms, feature engineering, interpretability, integration of multi-source data, system limitations, and recent advances primarily within the last five years.

Data Sources and Datasets

Heart disease prediction studies typically rely on structured clinical datasets encompassing patient demographics, risk factors, and clinical test results. A commonly used benchmark dataset consists of patient records containing attributes such as age, cholesterol levels, blood pressure, and ECG results. Additional datasets come from longitudinal cohort studies and health surveys, offering a wider range of clinical and lifestyle information. More recently, large-scale electronic

medical records (EMRs) and biobanks have been explored to incorporate a broader spectrum of patient data. Some approaches utilize natural language processing on clinical notes to extract relevant risk factors. Moreover, the advent of wearable sensor networks and Internet of Things (IoT) devices has enabled real-time physiological data collection, enriching prediction models with continuous heart rate, blood pressure, and oxygen saturation measurements.

A common challenge in heart disease datasets is class imbalance, where healthy cases substantially outnumber positive cases, potentially biasing models toward under-detection of disease. Researchers address this with sampling techniques such as oversampling minority classes or undersampling majority classes. Data quality issues, including missing or noisy entries, necessitate preprocessing strategies like imputation. Many datasets are relatively small and dated, highlighting the need for more contemporary, diverse, and larger-scale data to build robust predictive models.

Machine Learning Techniques

Various machine learning algorithms have been applied to heart disease prediction, with decision trees, logistic regression, support vector machines (SVM), k-nearest neighbors (KNN), random forests, Naïve Bayes, and ensemble methods being predominant.

- Decision Trees and Logistic Regression: These methods are favored for their interpretability and simplicity. Decision trees provide explicit decision rules, while logistic regression estimates the contribution of each feature linearly. However, standalone use may yield moderate accuracy compared to complex models, though they serve as effective baselines or components in ensemble systems.
- Support Vector Machines and k-NN: SVMs are powerful classifiers that maximize the margin between classes and handle non-linear data with kernel functions. They often achieve strong predictive performance but lack interpretability. KNN classifies based on similarity to neighbors and is valued for simplicity, commonly included in ensemble frameworks.
- Ensemble Methods: Combining multiple models is a successful strategy to improve prediction. Random forests, which aggregate many decision trees, are widely used due to their robustness and high accuracy. Gradient boosting techniques, such as XGBoost, further enhance performance by sequentially focusing on difficult samples. Stacking and voting ensembles, which integrate diverse classifiers, have demonstrated superior results, sometimes nearing perfect accuracy on curated datasets.
- Deep Learning and Hybrid Models: Deep neural networks, including convolutional and recurrent architectures, can learn hierarchical feature representations directly from raw data, such as ECG signals or images. Hybrid models combining deep learning with optimization algorithms have been developed to improve accuracy further. However, deep learning models require large datasets and substantial computational resources, and their opacity poses challenges for clinical acceptance.

Feature Selection and Optimization

Feature engineering plays a critical role in improving model performance. Methods such as mutual information ranking, recursive feature elimination, genetic algorithms, and regularization techniques are used to identify the most informative features and reduce redundancy. Hyperparameter tuning through grid search, Bayesian optimization, and evolutionary algorithms helps optimize model settings for better accuracy. Effective feature selection combined with robust model tuning enhances prediction and generalization capabilities.

Integration of Clinical, Behavioral, and Wearable Data

The field is moving toward integrating diverse data types beyond traditional clinical variables. Combining electronic

health records, genetic markers, laboratory tests, and physician notes provides a comprehensive patient profile. Lifestyle factors such as smoking, diet, physical activity, and stress are increasingly incorporated. Wearable technology enables continuous monitoring of physiological parameters, allowing dynamic and personalized risk assessment. Environmental data and behavioral patterns captured through smartphones and home sensors are emerging as additional context to refine predictions, supporting the trend toward holistic and precision medicine approaches.

Model Interpretability and Explainability

The "black box" nature of complex ML models limits their acceptance in healthcare. Explainable AI (XAI) methods, such as SHAP and LIME, have been employed to clarify model decisions by quantifying feature contributions. These explanations help clinicians understand and trust predictions, supporting clinical decision-making. Some approaches embed interpretability directly into model design or provide user-friendly interfaces for presenting risk assessments. Transparency fosters regulatory compliance and addresses liability concerns, which are crucial for clinical deployment.

Challenges and Limitations

Despite significant progress, heart disease prediction models face several challenges:

- Limited and outdated datasets constrain model generalization.
- Class imbalance can reduce sensitivity to disease cases.
- Overfitting is common, especially with small sample sizes.
- Missing and inconsistent data complicate feature use and comparison.
- Trade-offs exist between model accuracy and interpretability.
- Integration with clinical workflows and regulatory requirements remains difficult.

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

Future Directions

Future advancements rely on interdisciplinary collaboration among data scientists, clinicians, and ethicists. Incorporating clinical expertise in feature selection and model validation is essential. Federated learning offers privacy-preserving multi-institutional model training to improve generalizability. Multi-omics data integration and intrinsically interpretable models hold promise. Continuous monitoring via wearables and real-time data processing will enable dynamic risk prediction. Ethical considerations, fairness, and regulatory compliance must be embedded in system design to ensure equitable and safe healthcare delivery.

Table 2.1: literature survey

AUTHOR	ALGORITHM/TECHNIQUE	METHODOLOGY	REMARKS/PROBLEM	MERITS
1:Patel, Harshil 2:Singh, Rahul 3:Verma, Ankit April 2023	-Random Forest -Support Vector Machine (SVM) -Logistic Regression	Utilized the UCI Heart Disease dataset. Applied feature selection using correlation analysis, trained Random Forest, SVM, and Logistic Regression models to predict heart disease presence.	Random Forest achieved highest accuracy. Challenge: Handling imbalanced dataset and feature redundancy.	High accuracy Robust classification, Feature importance helps interpretability, Handles non-linear data well.
1:Sharma, Neha 2:Gupta, Rohit 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.	Effective integration of clinical and signal data. Requires large dataset and computational power.	Better generalization and automatic feature extraction from ECG, High prediction accuracy (~90%), Can handle multimodal data.

1:Patil, Sachin 2:Desai, Priya January 2024	-K-Nearest Neighbors (KNN) -Naive Bayes	Used patient demographic and clinical data. Trained KNN and Naive Bayes classifiers to predict heart disease risk. Evaluated with accuracy, precision, recall metrics.	KNN sensitive to feature scaling and choice of K. Naive Bayes faster but less accurate.	Simple algorithms, Good baseline performance, Useful for quick prototyping

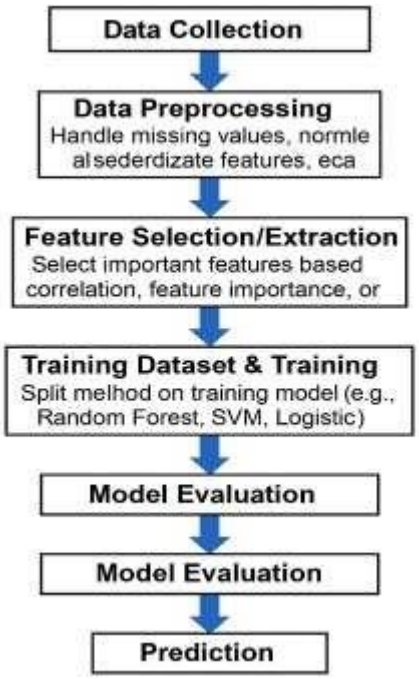

Figure 2.1: Existing block diagram for Heart Disease Prediction Using Machine Learning

2.1.1. **Data Collection**

This is the first step, where patient data is gathered. This could include medical history, test results, lifestyle data, etc.

2.1.2. **Data Preprocessing**

Preprocessing involves cleaning the data by handling **missing values**, **normalizing or standardizing** features, and possibly encoding categorical variables.

2.1.3. **Feature Selection/Extraction**

Important features are selected based on techniques like **correlation**, **feature importance**, or other statistical/ML-based

methods.

2.1.4. **Training Dataset & Training**

The data is split into **training and testing datasets**. Machine learning models like **Random Forest, SVM, or Logistic Regression** are trained on this data.

2.1.5. **Model Evaluation**

The model is tested using performance metrics (e.g., accuracy, precision, recall, F1 score) to evaluate how well it predicts heart disease.

2.1.6 **Prediction**

Based on the trained and evaluated model, predictions are made on new/unseen data to determine the likelihood of heart disease.

.Problem statement

Despite the structured architecture of machine learning models for heart disease prediction, several limitations hinder their effectiveness and broader applicability. A major concern lies in the reliance on traditional feature selection techniques, which may not adequately capture complex, nonlinear relationships among clinical variables such as cholesterol, blood pressure, and ECG readings. Additionally, many models fail to account for feature interactions or latent patterns in medical data, particularly when limited to shallow learning algorithms or basic statistical models.

Another challenge is the **class imbalance** commonly found in healthcare datasets, where positive heart disease cases are significantly outnumbered by negative cases, leading to biased model performance. Moreover, models trained on small or institution-specific datasets often suffer from **poor generalization** across diverse patient populations, increasing the risk of **overfitting**.

Interpretability is another crucial issue. Many machine learning models, particularly ensemble methods and neural networks, operate as **black boxes**, offering little insight into the decision-making process—an unacceptable limitation in clinical settings where transparency and trust are essential. The **absence of domain-informed explainability** reduces clinician adoption and confidence in AI-driven diagnostics.

Furthermore, **real-time deployment** is often restricted by computational complexity, and the presence of **missing or noisy medical data** further complicates accurate prediction. These limitations collectively emphasize the need for more robust, explainable, and clinically adaptable machine learning frameworks tailored for heart disease prediction.

Heart disease remains a leading cause of morbidity and mortality worldwide, necessitating the development of early and accurate predictive models. While machine learning (ML) offers promising capabilities in this regard, several critical challenges persist that affect the reliability, usability, and clinical integration of such models.

One significant issue is the **quality and diversity of clinical datasets**. Publicly available datasets such as the UCI Heart Disease dataset are relatively small, lack demographic diversity, and are often outdated. This results in **poor generalizability** across different populations, ethnicities, and clinical environments. Furthermore, these datasets often contain **missing values**, **class imbalance**, and **redundant or irrelevant features**, all of which can distort model performance and result in misleading outcomes.

Another key limitation lies in the **choice of machine learning algorithms**. Many studies rely on conventional classifiers

such as Logistic Regression, Support Vector Machines, or Decision Trees, which may lack the expressive power to capture complex, non-linear interactions among biomedical variables. Deep learning methods such as Artificial Neural Networks (ANNs) can address this to an extent, but they introduce new concerns such as **increased computational cost**, **overfitting**, and **reduced interpretability**.

Moreover, most existing models do not incorporate **temporal patterns** in patient health data, such as changes in ECG or blood pressure over time. The **lack of time-series analysis** prevents models from leveraging trends and progression indicators, which are often crucial in detecting the onset of cardiovascular diseases.

proposed block diagram

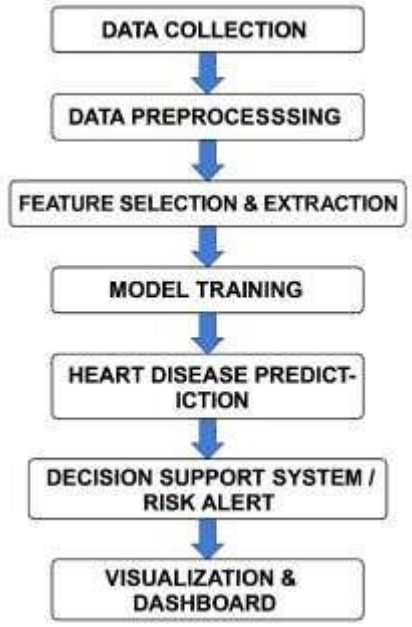

Figure 2.2: Heart disease prediction using Machine Learning

2.2 Software and Tools used :

2.2.1. Python

2.3.2. Visual Studio Code (VS Code)Python is the primary programming language used in this project due to its simplicity, rich ecosystem of data science libraries, and seamless integration with machine learning workflows. Its robust set of libraries such as **Pandas**, **NumPy**, **Scikit-learn**, and **Streamlit** makes it highly suitable for developing predictive models and deploying them as interactive applications.

In this project, Python facilitates the end-to-end machine learning pipeline—from **data preprocessing** and **feature scaling** to **model training**, **evaluation**, and **real-time prediction**. The **Scikit-learn** library provides efficient implementations of key algorithms such as **Random Forest Classifier** and **Logistic Regression**, enabling reliable classification of heart disease risk based on patient attributes.

Moreover, **Streamlit**, a powerful Python-based web framework, is utilized to build an intuitive and interactive user interface. It allows users to input medical parameters, get instant predictions, and visualize results without needing to understand the underlying code.

Python's modular and readable syntax accelerates development and debugging, ensuring that both technical users and healthcare stakeholders can benefit from the application effectively. This combination of powerful libraries and development ease makes Python the ideal choice for implementing machine learning-based health diagnostics.

2.3.2 Visual Studio Code (VS Code)

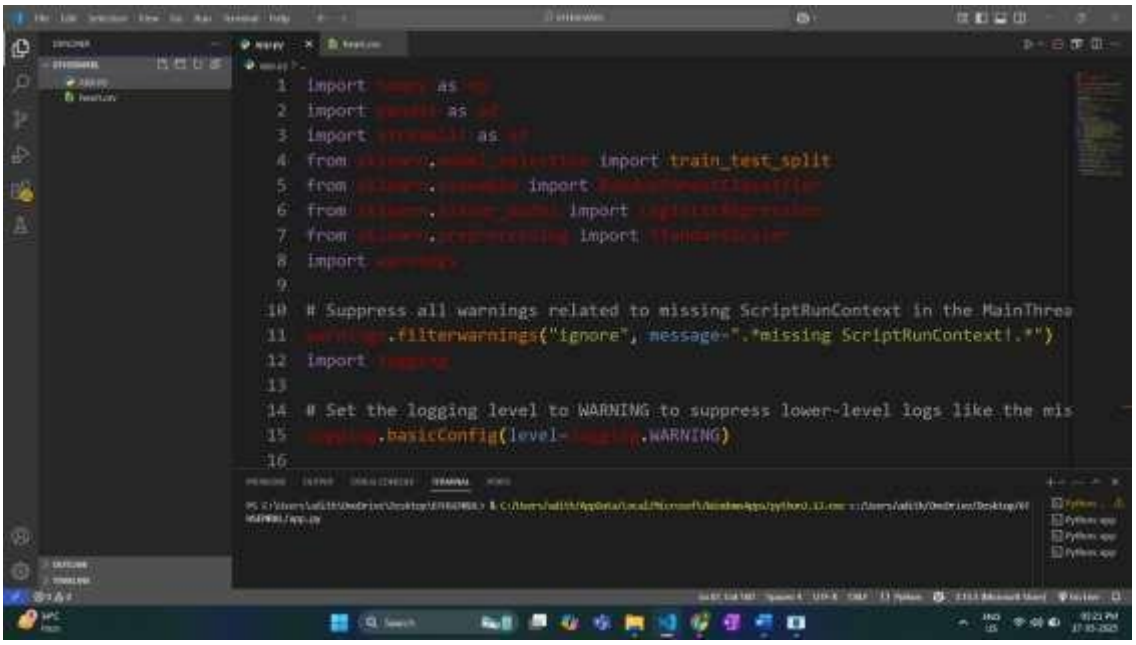

Figure 2.3: Visual Studio Code

In my **Heart 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
- Performs well on classification tasks like 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 "No Heart Disease").

- **Why it's used:**
  - Simple and interpretable
  - Good baseline model
  - Works well when there's a linear relationship between features and output

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 Scikit-learn, 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

PowerShell or the Command Prompt is used to execute Python scripts, activate the virtual environment, and install required dependencies using pip. These terminals play an essential role in managing the runtime execution of your application, especially when running scripts like train.py, predict.py, or activating your venv. They also help track error logs and outputs during real-time model predictions.

2.3.5. NumPy

NumPy (Numerical Python) is a fundamental library for numerical computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a wide collection of mathematical functions to operate on these arrays. In this project, NumPy is primarily used to format user inputs into structured arrays that can be passed to machine learning models for prediction. It ensures efficient numerical computations, which is especially important when dealing with real-time data inputs

2.3.6. Pandas

Pandas is a powerful library for data manipulation and analysis. It offers data structures like DataFrame and Series that simplify tasks such as reading, cleaning, and transforming datasets. In this project, Pandas is used to load the heart disease dataset from a CSV file and prepare it for model training by separating the features and target variable. It provides easy-to-use functions to inspect and preprocess the data before feeding it into machine learning models.

2.3.7. 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.

2.3.8. Sklearn.model_selection

a. sklearn.model_selection

This module provides tools for splitting datasets into training and testing subsets. In this project, the train_test_split() function is used to divide the data into 80% training and 20% testing, helping to evaluate the model's performance on unseen data.

b. sklearn.ensemble.RandomForestClassifier

This submodule contains the implementation of the Random Forest Classifier, an ensemble learning method that builds

multiple decision trees and combines their outputs to improve accuracy. It is one of the primary models used for predicting the presence of heart disease based on medical features.

c. sklearn.linear_model.LogisticRegression

This module provides the Logistic Regression model, which is a statistical method for binary classification. It is used in the project as a secondary prediction model to compare with the Random Forest classifier. Logistic regression works well for datasets where the outcome is categorical (such as "disease" or "no disease").

d. sklearn.preprocessing.StandardScaler

This module is used for feature scaling. The StandardScaler standardizes features by removing the mean and scaling to unit variance. This is crucial for models like Logistic Regression, which perform better when input features are on the same scale.

2.3.4. 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.

2.3 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

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)

3 Implentation

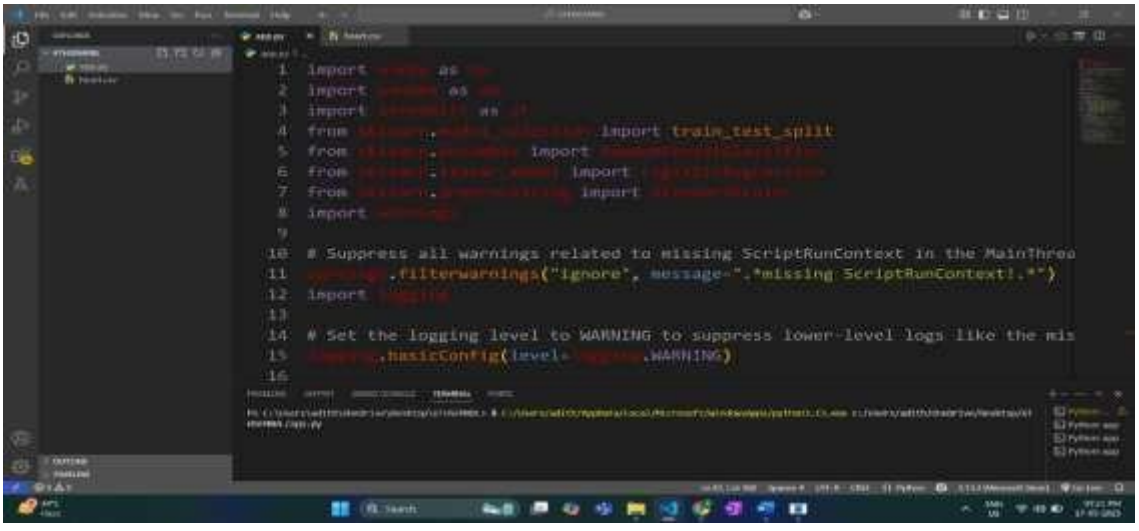

Figure 3.1: train model

Steps for implementation

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

2 Set Up a Virtual Environment

3 Install Dependencies(tensorflow keras opencv-python matplotlib pandas)

4 Download & Preprocess the Dataset

4 Results and discussion

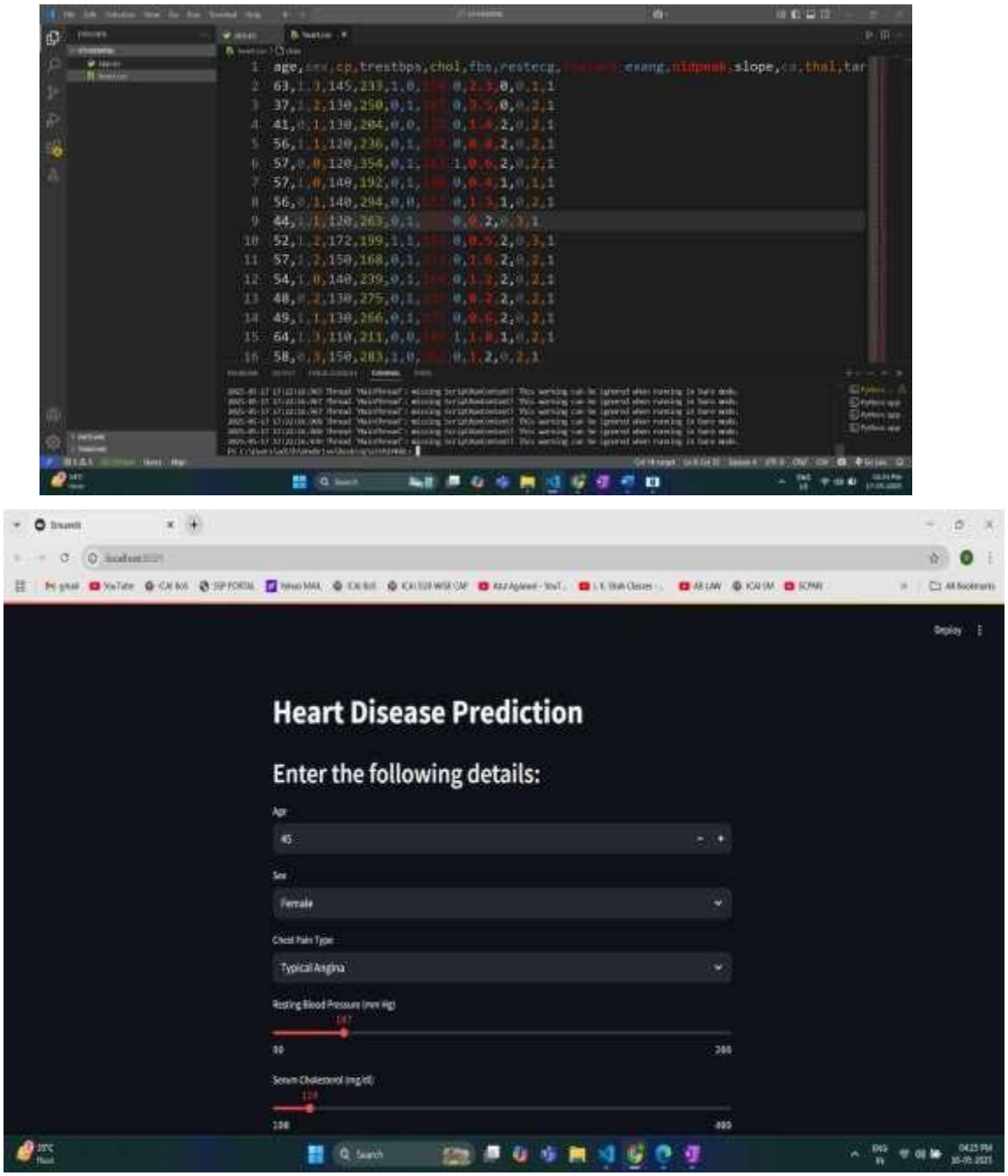

Figure 4.1: Dataset path

Figure 4.2: Model path

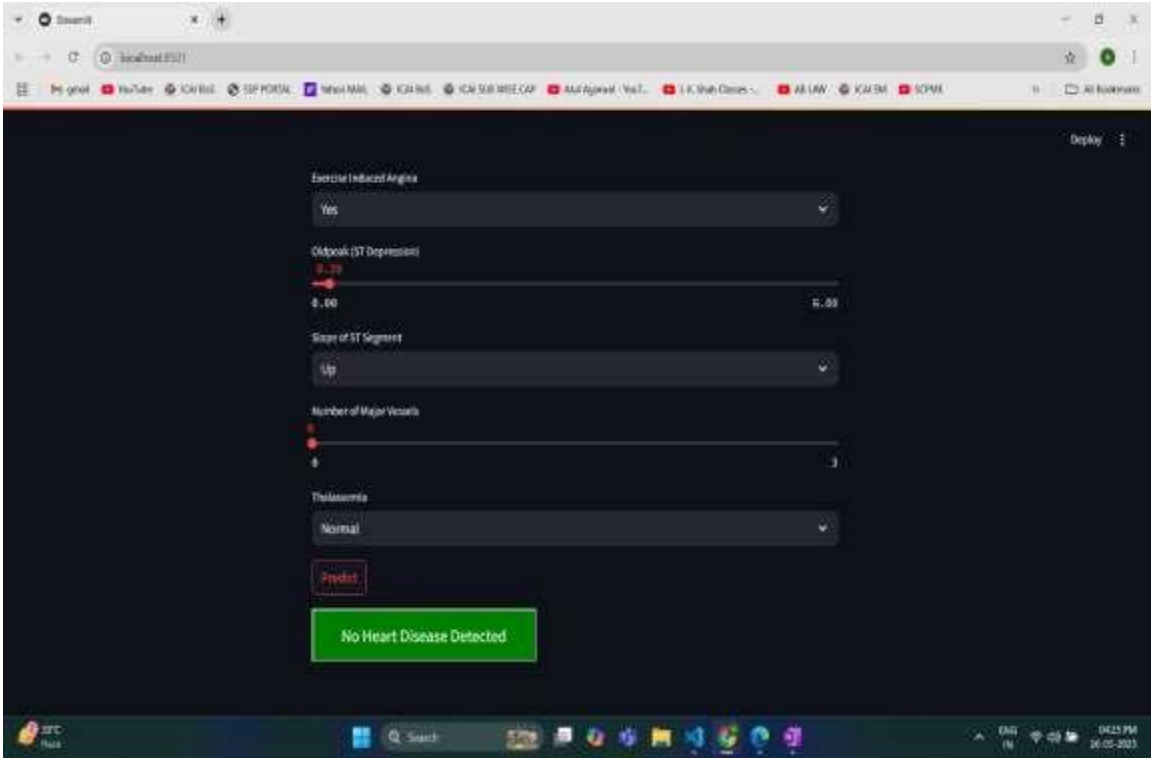

Figure 4.3: Output of no heart disease prediction

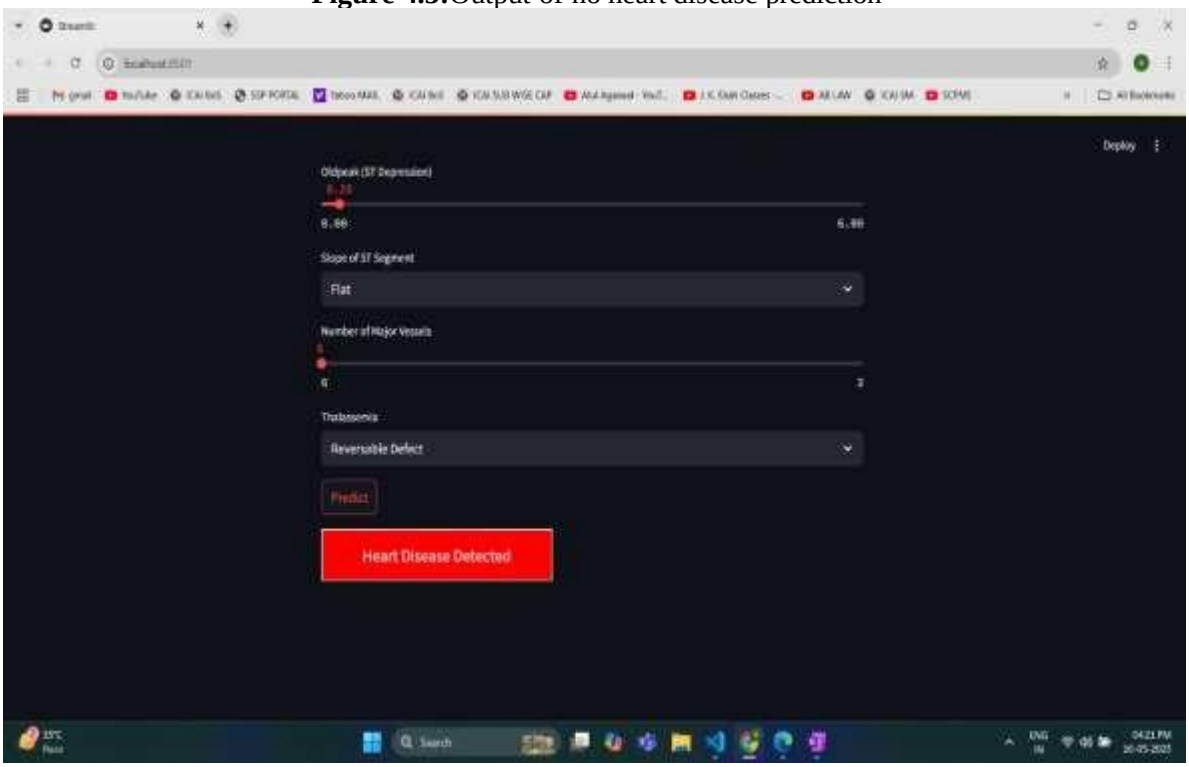

Figure 4.4: Output of heart disease prediction

Table no 4.1;dgsfh

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 4.2;dgsfh

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

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 4.4;dgs

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
18	58	0	2	120	340	0	1	172	0	0	2	0	2	1
19	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
20	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
21	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
22	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1

Table no 4.5;dgsfh

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
18	58	0	2	120	340	0	1	172	0	0	2	0	2	1
19	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
20	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
21	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
22	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1

5 conclusion and future scope Author credentials

Tammali Adithya Vardhan 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 Internet of Things

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