Visvesvaraya Technological University Belagavi, Karnataka

R.T.E Society's

Rural Engineering College Hulkoti – 582205

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Mini Project Report On

"CARDIOVASCULAR DISEASE PREDICTION USING MACHINE LEARNING"

Under the Guidance of

Prof. NEHABANU.H

Project Associates:

ROSHANI S DAFEDAR	2RH21CS038
AZAM PASHA	2RH21CS010
SHIVANAND MADIWALAR	2RH21CS044
VINAYAK HIREMATH	2RH21CS060

R.T.E Society’s Rural Engineering College , Hulkoti – 582205

Department of Computer Science & Engineering

2023-2024

CERTIFICATE

Certified that mini project work entitled

“CARDIOVASCULAR DISEASE PREDICTION USING MACHINE LEARNING”

Carried out by Miss. **Roshani S Dafedar**(USN 2RH21CS038), Mr.**Azam Pasha**(2RH21CS010),Mr.**Shivanand Madiwalar**(2RH21CS044)AND Mr**.Vinayak Hiremath** (2RH21CS060) are bonafide students of Department of Computer Science and Engineering, in partial fulfillment for the award of Degree of Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belagavi during the year 2023-24. It is certified that all corrections/suggestions indicated for internal assessments have been incorporated in the report deposited in the departmental library. The mini project report has been approved as it satisfies the academic requirements in respect of mini project work prescribed for the said degree.

Signature of the Guide	**Signature of the HOD**	**Signature of the Principal**
Prof. Nehabanu.H	Dr.S.H Angadi	Dr.V.M. Patil

ACKNOWLEDGEMENT

The knowledge and satisfaction that accompanies a successful completion of a project phase1 is hard to describe. Behind any successful project there are wise people guiding throughout. We thank them for guiding us, correcting our mistakes, and providing valuable feedback. We would consider it as our privilege to express our gratitude and respect to all those who guided and encouraged us in this project.I would like to thank our Dean Academics **Dr.L.R.Somangoudar** for his constant support and guidance.We extend our heartfelt gratitude to our beloved principal **Dr. V. M. Patil**, REC Hulkoti, for the success of this project.

We are grateful to **Dr. S H Angadi**, Head of CSE Department Rural Engineering College Hulkoti, for providing support and encouragement.

We convey our sincerest regards to our project guide, **Prof.Nehabanu.H** And our co-ordinator **Prof.M.V Umarani** , Dept. of CSE, Rural Engineering College Hulkoti, for providing guidance and encouragement at all times needed.

PROJECT ASSOCIATES

ROSHANI S DAFEDAR 2RH21CS038
AZAM PASHA 2RH21CS010
SHIVANAND MADIWALAR 2RH21CS044
VINAYAK HIREMATH 2RH21CS060

ABSTRACT

Cardiovascular diseases (CVD) are a leading cause of mortality worldwide, necessitating effective prediction models to identify at-risk individuals early. This project aims to develop a machine learning model using the k-Nearest Neighbors (kNN) algorithm to predict the likelihood of cardiovascular disease in patients based on clinical and demographic data. The dataset is sourced from Kaggle and undergoes several preprocessing steps, including handling missing values, normalization, and scaling of features. The kNN algorithm is then employed to build the prediction model. The model's performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1.About CVD

Cardiovascular diseases (CVDs) remain the leading cause of mortality globally, accounting for approximately 31% of all deaths. Early detection and prevention are crucial to reduce the burden of these diseases, enabling timely intervention and treatment. Traditional methods of diagnosis often involve extensive medical tests and expert evaluation, which can be time-consuming and resource-intensive. In recent years, advancements in machine learning (ML) have provided new avenues for the development of predictive models that can assist in the early detection of CVDs.

This project focuses on utilizing the k-Nearest Neighbors (kNN) algorithm, a simple yet effective machine learning technique, to predict the likelihood of cardiovascular disease in individuals based on a set of clinical and demographic parameters. The kNN algorithm is chosen for its simplicity and effectiveness in classification tasks, making it suitable for this application.

The dataset used in this project is sourced from Kaggle, a popular platform for data science and machine learning competitions. It contains a variety of health indicators that are commonly associated with cardiovascular conditions, such as age, sex, chest pain type, resting blood pressure, serum cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, old peak (ST depression induced by exercise), the slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy, and thalassemia.

The project is structured into several key phases: data preprocessing, model training, evaluation, and deployment. Data preprocessing involves cleaning the dataset, handling missing values, and normalizing features to ensure the data is suitable for model training. The kNN algorithm is then used to train the model on the preprocessed data, and its performance is evaluated using metrics such as accuracy, precision, recall, and F1-score.

1.2.Types of cardiovascular diseases

1. **Coronary Artery Disease (CAD)**: This is the most common type of heart disease, caused by the buildup of plaque in the coronary arteries, which supply blood to the heart muscle. It can lead to chest pain (angina), heart attacks, and heart failure.
2. **Heart Attack (Myocardial Infarction)**: Occurs when the blood flow to a part of the heart muscle is blocked, usually by a blood clot, causing damage to the heart muscle.
3. **Heart Failure**: A condition where the heart is unable to pump blood effectively, leading to symptoms like shortness of breath, fatigue, and fluid retention.
4. **Arrhythmias**: Abnormal heart rhythms that can be too fast, too slow, or irregular. Common types include atrial fibrillation, ventricular tachycardia, and bradycardia.
5. **Congenital Heart Disease**: Refers to heart abnormalities present at birth. These can involve the heart valves, chambers, or blood vessels and vary in severity.
6. **Cardiomyopathy**: A disease of the heart muscle that makes it harder for the heart to pump blood. Types include dilated, hypertrophic, and restrictive cardiomyopathy.
7. **Peripheral Artery Disease (PAD)**: A condition where the arteries that supply blood to the limbs are narrowed, leading to pain and circulation issues, usually in the legs.
8. **Aortic Aneurysm**: An abnormal bulge in the wall of the aorta, which can rupture and cause life-threatening bleeding.
9. **Heart Valve Disease**: Involves damage to or a defect in one of the four heart valves: mitral, aortic, tricuspid, or pulmonary. This can lead to regurgitation, stenosis, or atresia.
10. **Stroke**: Occurs when the blood supply to part of the brain is interrupted or reduced, preventing brain tissue from getting oxygen and nutrients. It can be ischemic (caused by a blockage) or hemorrhagic (caused by bleeding).
11. **Hypertension (High Blood Pressure)**: A chronic condition where the force of the blood against the artery walls is too high, which can lead to other CVDs.
12. **Rheumatic Heart Disease**: Caused by rheumatic fever, which can damage the heart valves and lead to heart failure and other complications.

These diseases often share risk factors such as high blood pressure, high cholesterol, smoking, diabetes, obesity, physical inactivity, and poor diet. Early detection and management are crucial to prevent complications and improve outcomes for individuals with cardiovascular diseases.

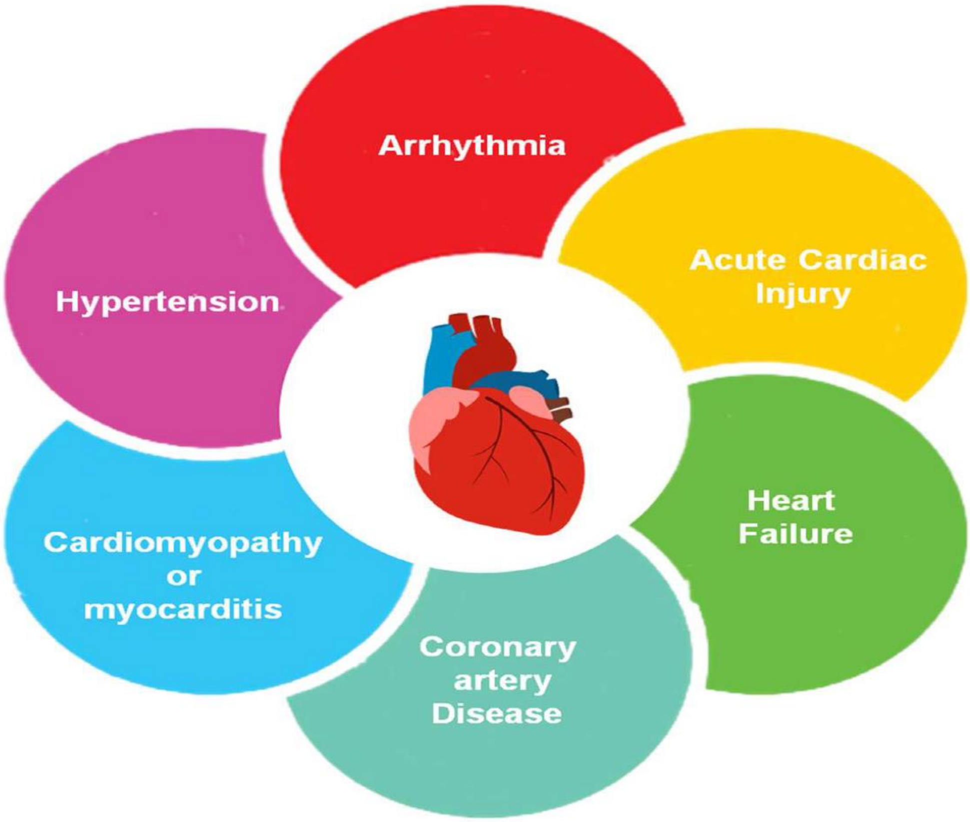

Fig 1.2.1:Types of Cardiovascular Diseases

Overall, this project demonstrates the potential of machine learning in transforming healthcare by providing tools for early disease detection and prevention. By leveraging the kNN algorithm and an intuitive GUI, this project aims to contribute to the ongoing efforts in improving cardiovascular health outcomes through innovative technological solutions.

CHAPTER 2

LITERATURE SURVEY

1. Senthilkumar Mohan .et.al *"**Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques"*** , they proposed a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. Then produce an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM).

2. Archana Singh .et.al "***Heart Disease Prediction Using Machine Learning Algorithms*"** They calculated the accuracy of machine learning algorithms for predicting heart disease, for this algorithms are k-nearest neighbor, decision tree, linear regression and support vector machine(SVM) by using UCI repository dataset for training and testing. For implementation of Python programming Anaconda(jupytor) notebook is best tool, which have many type of library, header file, that make the work more accurate and precise.

3. Vijeta Sharma .et.al "***Heart Disease Prediction using Machine Learning Techniques*"** The objective of this paper is to build a ML model for heart disease prediction based on the related parameters. They have used a benchmark dataset of UCI Heart disease prediction for this research work, which consist of 14 different parameters related to Heart Disease. Machine Learning algorithms such as Random Forest, Support Vector Machine (SVM), Naive Bayes and Decision tree have been used for the development of model.

4. Dimitris Bertsimas .et.al "***Machine Learning for Real-Time Heart Disease Prediction*" ,** they proposed a novel methodology to extract ECG-related features and predict the type of ECG recorded in real time (less than 30 milliseconds). Those models leverage a collection of almost 40 thousand ECGs labeled by expert cardiologists across different hospitals and countries, and are able to detect 7 types of signals: Normal, AF, Tachycardia, Bradycardia, Arrhythmia, Other or Noisy. We exploit the XGBoost algorithm, a leading machine learning method, to train models achieving out of sample F1 Scores in the range 0.93 – 0.99.

5. Tülay KarayÕlan .et.al "***Heart Disease Using Neural Network***". In this paper, a heart disease prediction system which uses artificial neural network backpropagation algorithm is proposed. 13 clinical features were used as input for the neural network and then the neural network was trained with backpropagation algorithm to predict absence or presence of heart disease with accuracy of 95%.

6. P. Ramprakash .et.al **"*Heart Disease Prediction Using Deep Neural Network*"** In this paper- They proposed a model that shows better results on both the testing and training data. DNN and ANN were used to analyse the efficiency of the model which accurately predicts the presence or absence of heart disease.

7. Rahul Katarya .et.al "***Predicting Heart Disease at Early Stages using Machine Learning: A Survey"*** In this paper- Supervised machine learning techniques used in this prediction of heart disease are artificial neural network (ANN), decision tree (DT), random forest (RF), support vector machine (SVM), naïve Bayes) (NB) and knearest neighbour algorithms gave around 93% accuracy.

8. Harshit Jindal .et.al *"**Heart disease prediction using machine learning algorithms"*** The research paper mainly focuses on which patient is more likely to have a heart disease based on various medical attributes. They prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. This project gives us significant knowledge that can help us predict the patients with heart disease It is implemented on the .pynb format.

From the above survey these studies collectively illustrate the promising role of machine learning and neural networks in heart disease prediction. The high accuracy rates achieved by various models underscore the potential of these technologies to significantly impact healthcare by enabling early detection and intervention, ultimately contributing to better patient outcomes and reducing the burden of heart disease.

CHAPTER 3

PROPOSED MODEL

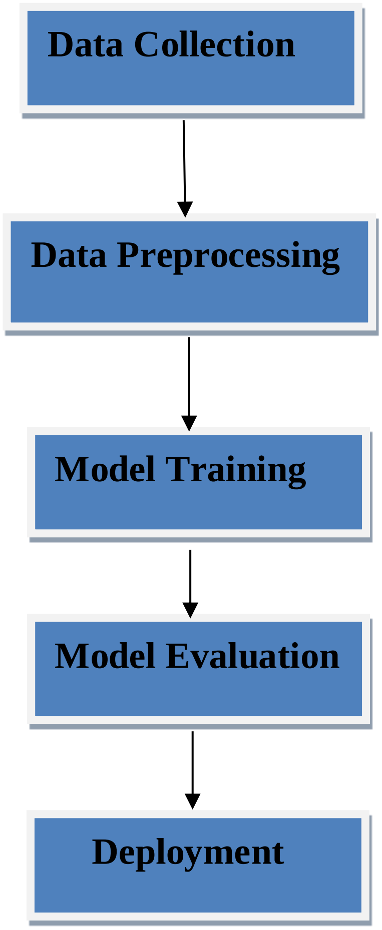

Fig 3.1:Block diagram of the process

3.1.Problem Statement:

This project aims to develop a machine learning-based solution for predicting cardiovascular diseases using the k-Nearest Neighbors (kNN) algorithm. The goal is to create an accurate, efficient, and user-friendly tool that can assist healthcare professionals in identifying patients at risk of CVD, enabling timely and appropriate medical responses.

3.2.Objectives:

- **Develop a Predictive Model:**

- Utilize the k-Nearest Neighbors (kNN) algorithm to create a model that can accurately predict the presence of cardiovascular disease based on patient data.

- **Data Preprocessing:**

- Implement data preprocessing techniques to clean, normalize, and prepare the dataset for optimal performance of the kNN algorithm.

- **User Interface Development:**

- Create a graphical user interface (GUI) using Tkinter to allow users to input patient data and receive predictions in a user-friendly manner.

- **Model Evaluation:**

- Evaluate the performance of the kNN model using metrics such as accuracy, precision, recall, and F1-score to ensure its reliability and effectiveness.

- **Deployment and Accessibility:**

- Ensure the application is easily accessible and can be used by healthcare professionals with minimal technical expertise to facilitate early diagnosis and treatment planning.

- **Documentation:**

- Provide comprehensive documentation for the project, including an abstract, introduction, methodology, results, and conclusions to support further research and development.

- **Scalability and Flexibility:**

- Design the system to be scalable and flexible, allowing for future enhancements and integration of additional features or alternative machine learning algorithms.

- **Testing and Validation:**

- Conduct extensive testing and validation of the system with real-world patient data to confirm its accuracy and robustness in various clinical scenarios.

The proposed model for cardiovascular disease prediction leverages the k-Nearest Neighbors (kNN) algorithm to classify individuals as either at risk or not at risk for cardiovascular disease based on their clinical and demographic data. The model's workflow is organized into several stages: data collection, data preprocessing, model training, evaluation, and deployment. The following sections outline the proposed model in detail.

The k-Nearest Neighbors (kNN) algorithm is a simple yet powerful supervised learning technique used for both classification and regression tasks. It is one of the most intuitive machine learning algorithms, relying on the concept of similarity between data points. Below is an in-depth explanation of the kNN method.

Concept

The kNN algorithm classifies a data point based on how its neighbors are classified. It assumes that similar data points exist in close proximity. Therefore, if a majority of an instance's 'k' nearest neighbors belong to a particular class, the instance is assigned to that class.

3.3.How kNN Works

1. **Choosing the Number of Neighbors (k):**

- The user specifies 'k', the number of nearest neighbors to consider.
- The choice of 'k' affects the model's performance. A smaller 'k' might capture noise in the data, while a larger 'k' could smooth out the decision boundary.

2.**Calculating Distance:**

- The algorithm calculates the distance between the input data point and all points in the training dataset.
- Common distance metrics include Euclidean distance, Manhattan distance, and Minkowski distance.

- **Euclidean Distance:** ∑i=1n(xi−yi)2\sqrt{\sum_{i=1}^n (x_i - y_i)^2}∑i=1n(xi−yi)2
- **Manhattan Distance:** ∑i=1n∣xi−yi∣\sum_{i=1}^n |x_i - y_i|∑i=1n∣xi−yi∣

2. **Identifying Neighbors:**
   - The algorithm identifies the 'k' data points in the training set that are closest to the input data point.
3. **Voting for Classification:**
   - For classification, the algorithm assigns the class that is most common among the 'k' nearest neighbors.
   - For regression, the algorithm calculates the average of the values of the 'k' nearest neighbors.

Example of kNN in Classification

Suppose we have a dataset with two features, X and Y, and we want to classify a new point (A) as either class 0 or class 1.

1. **Select k (e.g., k = 3).**
2. **Calculate the distance from A to all other points in the dataset.**
3. **Identify the 3 closest points to A.**
4. **Assign the class based on the majority class of these 3 points.**

Strengths of kNN

- **Simplicity:** The algorithm is easy to understand and implement.
- **No Training Phase:** kNN is a lazy learner; it doesn't require a training phase. This can be advantageous for small datasets.
- **Adaptability:** kNN can be used for both classification and regression tasks.

Weaknesses of kNN

- **Computational Cost:** For large datasets, the algorithm becomes computationally expensive as it requires calculating the distance to all other points.
- **Storage Requirements:** Since kNN stores all training data, it requires a significant amount of memory.
- **Curse of Dimensionality:** The algorithm's performance can degrade in high-dimensional spaces because the distance metric becomes less informative.
- **Sensitivity to Irrelevant Features:** If the dataset contains irrelevant or redundant features, they can affect the distance calculations and, consequently, the model's performance.

Applications of kNN

- **Recommendation Systems:** Suggesting products or content based on the similarity to other users' preferences.
- **Medical Diagnosis:** Classifying medical conditions based on patient symptoms and historical data.
- **Pattern Recognition:** Identifying patterns in data, such as handwriting recognition.

CHAPTER 4

SYSTEM SPECIFICATION REQIREMENTS

4.1. Hardware Requirements

1. **Processor:**
   - Minimum: Dual-core processor
   - Recommended: Quad-core processor or higher
2. **Memory (RAM):**
   - Minimum: 4 GB
   - Recommended: 8 GB or higher
3. **Storage:**
   - Minimum: 10 GB of free disk space
   - Recommended: Solid State Drive (SSD) for faster read/write speeds
4. **Graphics:**
   - Integrated graphics card (sufficient for this project, no need for a dedicated GPU)

4.2.Software Requirements

1. **Operating System:**
   - Windows 10 or later
   - macOS 10.14 (Mojave) or later
   - Linux (Ubuntu 18.04 or later is recommended)
2. **Python:**
   - Python 3.6 or higher (Python 3.8 or 3.9 is recommended)
3. **Python Packages:**
   - Install necessary packages using pip:

Code:

```
pip install numpy pandas scikit-learn matplotlib joblib tkinter
```

4. **Integrated Development Environment (IDE):**

- Visual Studio Code (VS Code) is recommended.
- Other options: PyCharm, Jupyter Notebook, or any text editor of your choice.

5. **Version Control System:**

- Git (for version control and collaboration)

6. **Libraries and Tools:**

- **NumPy:** For numerical computations
- **Pandas:** For data manipulation and analysis
- **Scikit-learn:** For implementing the kNN algorithm and other machine learning tasks
- **Matplotlib:** For data visualization
- **Joblib:** For model serialization and deserialization
- **Tkinter:** For creating the graphical user interface (GUI)

CHAPTER 5

PROCEDURE-WISE DETAIL

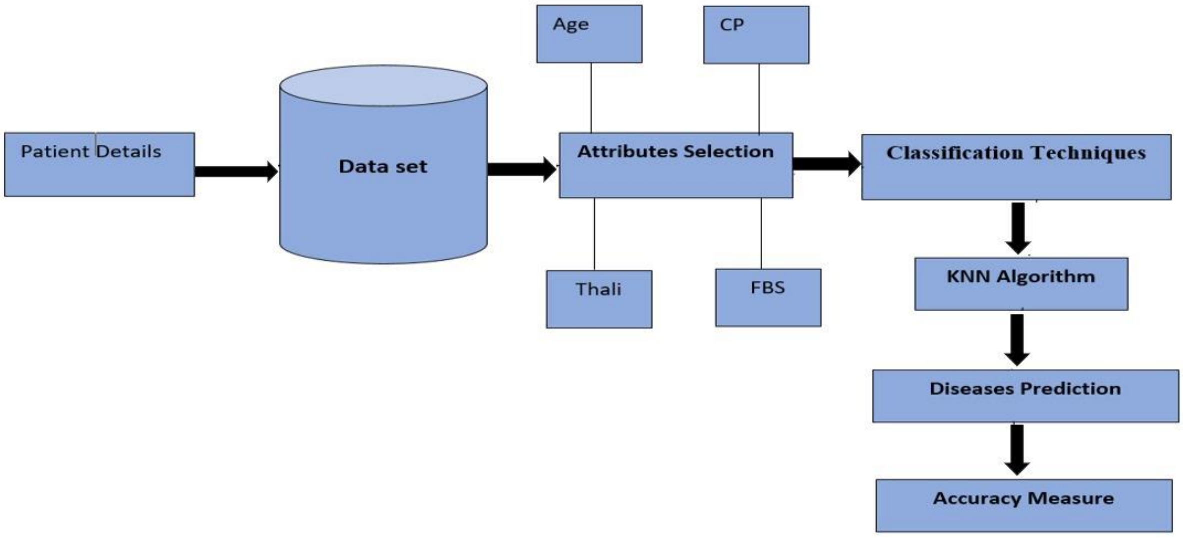

Fig 5.1:Architecture of Prediction Sysytem

1. Data Collection

The dataset for this project is sourced from Kaggle and includes a comprehensive set of health parameters related to cardiovascular conditions. The dataset comprises features such as age, sex, chest pain type, resting blood pressure, serum cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, old peak (ST depression induced by exercise), the slope of the peak exercise ST segment, the number of major vessels colored by fluoroscopy, and thalassemia.

2. Data Preprocessing

Data preprocessing is a critical step to ensure the dataset is clean and suitable for training the kNN model. This stage includes:

- **Handling Missing Values:** Missing values in the dataset are addressed through imputation techniques such as mean, median, or mode imputation, depending on the nature of the missing data.
- **Normalization:** Features are normalized to ensure they are on a comparable scale, which is essential for the distance-based kNN algorithm. Techniques like Min-Max scaling or Standardization are applied.

- **Encoding Categorical Variables:** Categorical features, such as sex and chest pain type, are encoded into numerical values using techniques like one-hot encoding or label encoding.

3. Model Training

The k-Nearest Neighbors algorithm is employed for this classification task. The kNN algorithm works by identifying the 'k' nearest data points to a given input and classifying the input based on the majority class among its neighbors. Key steps in this stage include:

- **Choosing the Value of k:** The optimal value of 'k' is determined through cross-validation. Various values are tested to find the one that offers the best performance.
- **Distance Metric:** The Euclidean distance metric is used to measure the distance between data points.

5. Deployment

The trained model is integrated into a user-friendly graphical user interface (GUI) developed using Tkinter. The GUI allows users to input patient data and receive a prediction regarding their cardiovascular disease risk. Key components of the deployment phase include:

- **Input Form:** A form for users to enter patient data such as age, sex, chest pain type, etc.
- **Prediction Display:** A section where the model's prediction is displayed.
- **Model Loading:** The trained kNN model is loaded using joblib, ensuring it is readily available for making predictions.

4. Model Evaluation

The performance of the kNN model is evaluated using the following metrics:

- **Accuracy:** The overall correctness of the model.
- **Precision:** The proportion of true positive predictions among all positive predictions.
- **Recall:** The proportion of true positive predictions among all actual positives.
- **F1-Score:** The harmonic mean of precision and recall, providing a balance between the two.

CHAPTER 6

RESULTS AND SNAPSHOTS

6.1.Dataset

This section presents the findings from the cardiovascular disease prediction project using the k-Nearest Neighbors (kNN) algorithm. The results are based on the evaluation of the model's performance on the test dataset.

A1 age

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	52	1	0	125	212	0	1	168	0	1	2	2	3	0
3	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
4	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
5	61	1	0	148	203	0	1	161	0	0	2	1	3	0
6	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
7	58	0	0	100	248	0	0	122	0	1	1	0	2	1
8	58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
9	55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
10	46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
11	54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
12	71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
13	43	0	0	132	341	1	0	136	1	3	1	0	3	0
14	34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
15	51	1	0	140	298	0	1	122	1	4.2	1	3	3	0
16	52	1	0	128	204	1	1	156	1	1	1	0	0	0
17	34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
18	51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
19	54	1	0	124	266	0	0	109	1	2.2	1	1	3	0
20	50	0	1	120	244	0	1	162	0	1.1	2	0	2	1
21	58	1	2	140	211	1	0	165	0	0	2	0	2	1
22	60	1	2	140	185	0	0	155	0	3	1	0	2	0
23	67	0	0	106	223	0	1	142	0	0.3	2	2	2	1
24	45	1	0	104	208	0	0	148	1	3	1	0	2	1
25	63	0	2	135	252	0	0	172	0	0	2	0	2	1
26	42	0	2	120	209	0	1	173	0	0	1	0	2	1

Fig 6.1.1:Csv file of Dataset

1.We extracted the Dataset from https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data.

2.The Dataset has 14 attributes such as age,sex,bp…..etc

6.2.User Interface for Cardiovascular Disease Prediction Project

As part of the cardiovascular disease prediction project, a graphical user interface (GUI) was developed to facilitate user interaction with the machine learning model. The GUI allows users to input patient data, load existing patient data, and predict the likelihood of cardiovascular disease. Below is a detailed description of the interface and its functionalities.

Once the data is entered or loaded, clicking the "Predict" button will use the trained kNN model to predict the presence or absence of cardiovascular disease. The result is then displayed to the user.

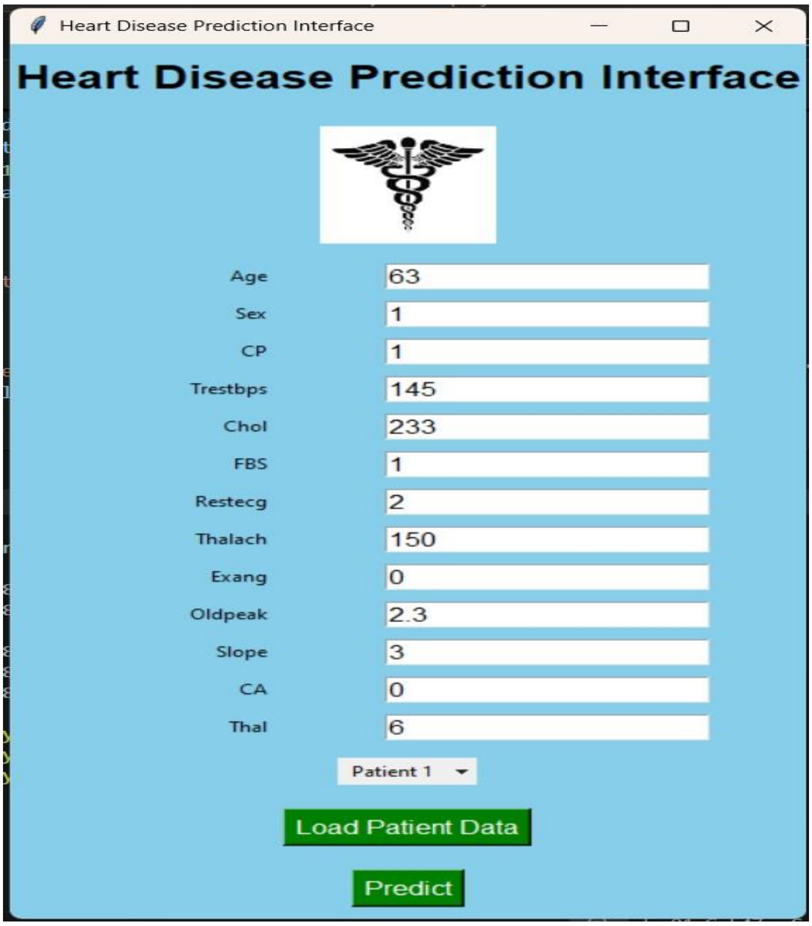

Fig 6.2.1:User interface to input patient details

- Once the patient details are inserted as input,the prediction result is displayed.
- The result is positive if the patient has heart disease.
- The result will show negative if the patient has no heart disease.

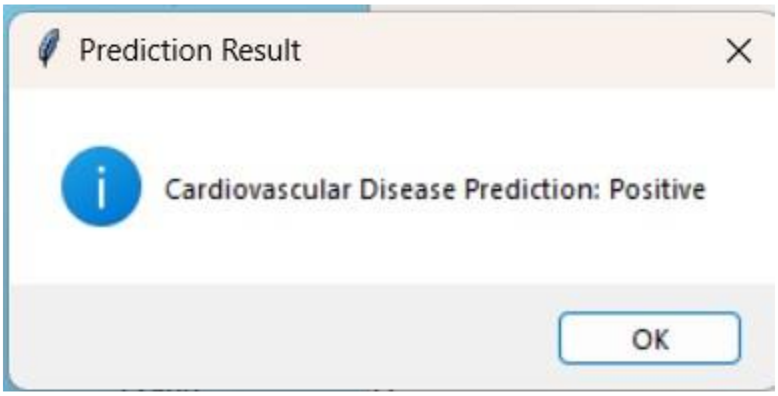

Fig 6.2.2:Prediction result(Positive)

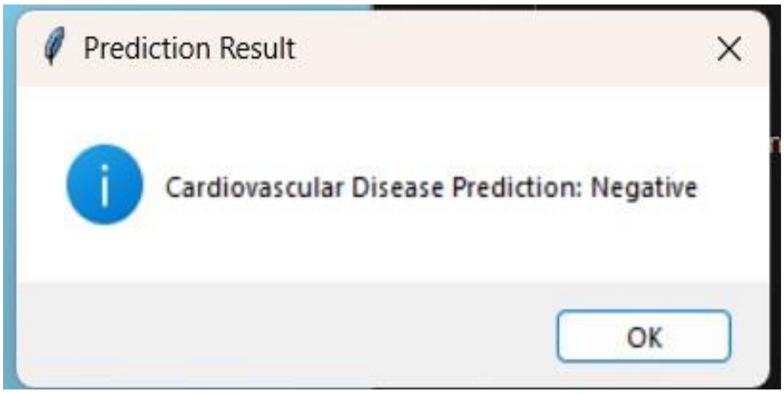

Fig 6.2.3:Prediction result(Negative)

6.3.Model Evaluation Metrics

The kNN model was evaluated using several performance metrics: accuracy, precision, recall, and F1-score. The following table summarizes the results:

```
TERMINAL
PS C:\Users\azamp\Desktop\Mini project> python -u "c:\Users\azamp\Desktop\Mini project\src\main.py"
Accuracy: 0.83
              precision    recall  f1-score   support

           0       0.88      0.77      0.82       102
           1       0.80      0.89      0.84       103

    accuracy                           0.83       205
   macro avg       0.84      0.83      0.83       205
weighted avg       0.84      0.83      0.83       205
```

Fig 6.3.1:The Accuracy Result in the terminal

Metric	Class 0 (No Disease)	Class 1 (Disease)	Average
Precision	0.88	0.80	0.84 (macro)
Recall	0.77	0.89	0.83 (macro)
F1-Score	0.82	0.84	0.83 (macro)
Support	102	103	205 (total)
Overall Accuracy			0.83 (83%)

Table 6.3.2:kNN performance Metrics

- **Precision:** The precision for class 0 (No Disease) is 0.88, indicating that 88% of the predictions for this class were correct. For class 1 (Disease), the precision is 0.80.
- **Recall:** The recall for class 0 is 0.77, showing that the model correctly identified 77% of the actual no disease cases. For class 1, the recall is 0.89.
- **F1-Score:** The F1-score for class 0 is 0.82, and for class 1, it is 0.84. The F1-score is the harmonic mean of precision and recall, providing a balance between the two.
- **Overall Accuracy:** The overall accuracy of the model is 83%, indicating that the model correctly predicted 83% of the instances in the test dataset.

6.4.Insights and Analysis

The evaluation metrics indicate that the kNN model performs well in predicting cardiovascular disease. Here are some key insights:

1. **Class Balance:** The dataset is relatively balanced, with 102 instances of no disease and 103 instances of disease. This balance helps in providing a fair evaluation of the model.
2. **Precision and Recall Trade-off:** The model demonstrates a good trade-off between precision and recall, with slightly higher recall for class 1 (disease), which is beneficial for medical diagnosis to minimize false negatives.
3. **F1-Score:** The F1-scores for both classes are comparable, indicating a balanced performance across both categories.

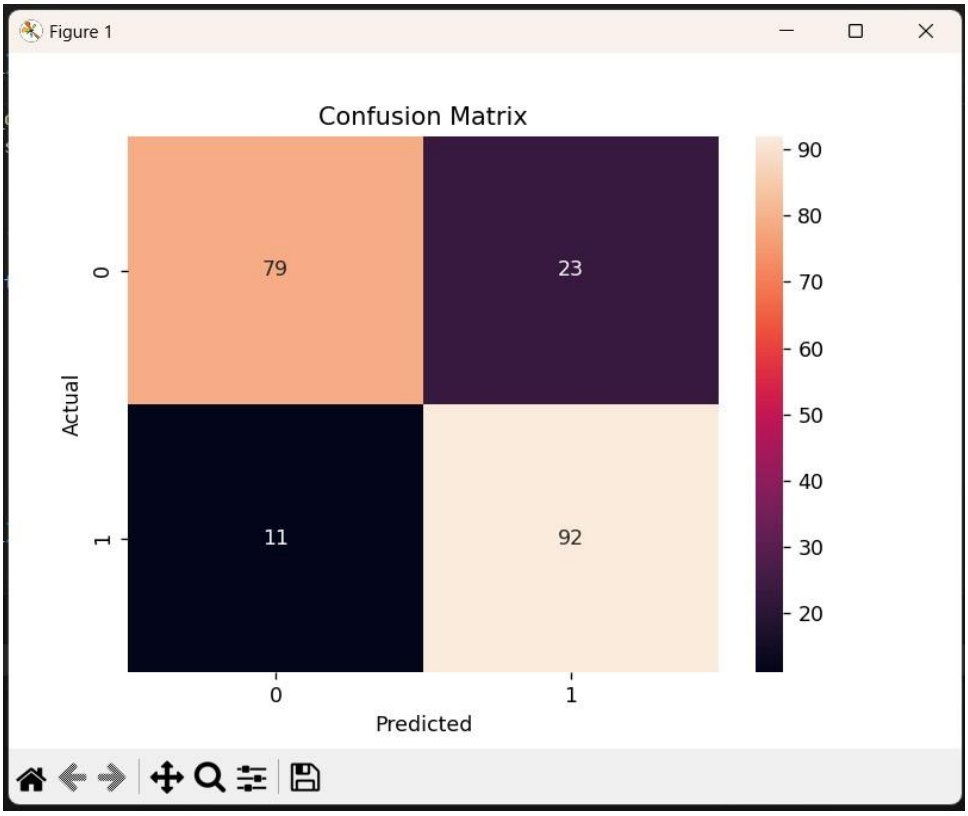

Fig 6.4.1:Confusion Matrix

A confusion matrix provides a detailed breakdown of the model's predictions:

	Predicted No Disease	**Predicted Disease**
Actual No Disease	79	23
Actual Disease	11	92

Table 6.4.2:Models Prediction

- **True Positives (TP):** 92 instances where the model correctly predicted disease.
- **True Negatives (TN):** 79 instances where the model correctly predicted no disease.
- **False Positives (FP):** 23 instances where the model incorrectly predicted disease.
- **False Negatives (FN):** 11 instances where the model incorrectly predicted no disease.

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

The cardiovascular disease prediction project utilizing the k-Nearest Neighbors (k-NN) algorithm has demonstrated its potential in identifying individuals at risk of cardiovascular conditions with reasonable accuracy. By leveraging historical health data, including features such as age, blood pressure, cholesterol levels, and other relevant metrics, the k-NN algorithm effectively classifies patients into risk categories based on their similarity to existing cases. The project's success highlights the capability of k-NN in handling complex, high-dimensional data and providing actionable insights for early intervention. This approach not only supports preventive healthcare measures but also emphasizes the value of machine learning in enhancing diagnostic accuracy and patient care. We have obtained the accuracy of 83%.

Experimenting with some machine learning algorithms, such as Random Forests or Support Vector Machines, and comparing their performance against k-NN might offer more robust predictions. Additionally, implementing cross-validation techniques and hyperparameter tuning could optimize model accuracy and reduce overfitting. Finally, integrating real-time data processing and developing a user-friendly interface for healthcare professionals could facilitate practical applications and broader adoption of the prediction system in clinical settings.

REFERENCES

1. Senthilkumar Mohan , Chandrasegar Thirumalai And Gautam Srivastava "***Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques"*** research paper published in the year 2019, School of Information Technology and Engineering, VIT University, Vellore 632015, India
2. Archana Singh And Rakesh Kumar "***Heart Disease Prediction Using Machine Learning Algorithms"*** research paper published in the year 2020, Madan Mohan Malaviya University of Technology Gorakhapur,India.
3. Dimitris Bertsimas, Luca Mingardi, and Bartolomeo Stellato ***"Machine Learning for Real-Time Heart Disease Prediction"*** research paper published this in the year 2020, bmitted on October 10th 2020 D.B. is with the Sloan School of Management, Massachusetts Institute of Technology.
4. Tülay KarayÕlan And Özkan KÕlÕç ***"Heart Disease Using Neural Network"*** research paper published in the year 2017, Department of Computer Engineering YÕldÕrÕm BeyazÕt University Ankara, Turkey.
5. P. Ramprakash ,R. Sarumathi, R. Mowriya and S. Nithyavishnupriya "***Heart Disease Prediction Using Deep Neural Network"*** research paper published in the year 2020.
6. Rahul Katarya and Polipireddy Srinivas "***Predicting Heart Disease at Early Stages using Machine Learning: A Survey"*** research paper published in the year 2020, Department of Computer Science and Engineering, Delhi Technological University, Delhi, India.
7. Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath ***"Heart disease prediction using machine learning algorithms"*** research paper published in the year 2020, Dept. Of Computer Science & Engineering Bharti Vidyapeeth's College Of Engineering, New Delhi.
8. Vijeta Sharma And Shrinkhala Yadav ***Heart Disease Prediction using Machine Learning Techniques*** research paper published in the year 2020, Computer Science, DST-Centre for Interdisciplinary Mathematical Science, Institute of Science, Banaras Hindu University Varanasi, India.
9. www.wikipedia.com