AI CARDIOLOGIST ADVANCEMENTS IN SUPERVISED LEARNING FOR HEART DISEASE PREDICTION

Mr. Ajith kumar[1], Mr. Srujan Kumar reddy[2], Mr. Ganga Raju[3], Ms. Revathy[4], Mr. Jaya Prakash[5], Ms. Chichu Nair[6]

Student[1,2,3], Assistant Professor[4], Professor[5,6]

Department of Computer Science and Engineering,

Dr. M.G.R. Educational And Research Institute

,Madurvoyal, Chennai-95,Tamil Nadu, India

Abstract

Heart disease is the biggest problem in the world and having no age to occur it can occur to a man or even a kid or any healthy person with the problems in their genes. Early and accurate prediction is crucial for effective preventive measures. This project explores the transformative potential of supervised learning within Artificial Intelligence (AI) for heart disease prediction. We used many algorithms like KNN and Tree algorithm to study more about the disease-predicting technology. The project examines how AI cardiologists, by using the algorithms can able to take the data of the user or a patient and can help the patient to get a review of the disease and the stage they were in, This type of analysis holds trust for more accurate and personalized risk assessments, leading that resulted in data to get a new outcome on how to prevent the problem. We understand all of the challenges of data quality, bias mitigation, and explainability in AI models, emphasizing the importance of ethical considerations in their development and deployment. Finally, the paper discusses the future directions of AI-powered heart disease prediction, exploring the potential of emerging techniques like explainable AI and federated learning to advance this field further.

Keywords

Cardiovascular disease (CVD) prediction, Supervised learning, Artificial Intelligence (AI), Heart disease, Algorithms (logistic regression, support vector machine)

Literature Survey

1. Gárate-Escamila et al. (2020)

This study explored the use of **“Deep Neural Networks (DNN)”** and **“Artificial Neural Networking (ANN)”** for predicting outcomes using an X2X^2X2 statistical model. The research utilized **clinical data parameters**, emphasizing their importance in training models to improve prediction accuracy. This approach highlights how combining statistical models with advanced neural networks can enhance the performance of medical predictions.

2. Harvard Medical School (2020)

Analyzers at Harvard Medical School utilized **Many datasets from war and brutal places** for analyzing cardio diseases from the data they gathered. They applied various machine learning classifiers to improve predictive capabilities. To enhance the model's performance and manage large feature spaces, **"Principal Component Analysis (PCA)"** was used for **"dimensionality reduction** and **feature selection"**. This study shows us the importance of dimensionality reduction techniques in medical diagnostics.

3. Zhang et al. (2018)

Zhang and colleagues combined the **AdaBoost classifier** with **PCA** for feature extraction, which significantly improved the accuracy of predictions. The integration of AdaBoost, a boosting algorithm, with PCA, shows how combining ensemble learning methods and dimensionality reduction can lead to better feature selection and robust results in predictive modeling.

4. Singh et al. (2018)

This study focused on detecting **coronary artery disease** using **heart rate variability (HRV)** as a key indicator. They employed **Fisher's method** and **generalized discriminant analysis** to select important features and binary classifiers for detection. This work highlights advanced statistical methods to extract relevant features from physiological data to improve diagnostic accuracy.

5. Yang and Nataliani (2018)

The researchers employed a **fuzzy clustering** method, specifically **fuzzy c-means**, to perform feature reduction. By assigning weighted importance to various features, this approach allowed the model to show all the interest in all possible places as it shows in the dataset. It will illustrate the potential of fuzzy logic in improving data preprocessing for machine learning tasks.

Introduction

Heart disease or cardiovascular disease (CVD) is a major health problem experienced across the globe. It continues to be one of the main factors in mortality and persistent effects on population health. CVD claims approximately 7,500 people daily, and in India; CVD alone contributes to 28.1% of all reported deaths. This paper aims to establish that early and accurate heart disease diagnosis is crucial in the management of diseases in a bid to minimize complications and maximize patients' survival. Researchers based on data science and machine learning ideas have tried to see the possibility of using numerous algorithms and datasets to predict the risk of heart diseases incrementally. In this project, we focus on performing exploratory data analysis on a given data set with a large number of parameters related to heart attack and cardiovascular diseases. The main goal of this work is to create a model that can predict individuals who may develop heart diseases. To this end, only visualization and data analysis methods will be used for analyzing the relationships between attributes involved to obtain insightful results regarding the state of heart health. One of the most critical stages of our research process will be the finding of the correlation between different features, which will help to determine which of them should be taken into account to build the predictive model. The main place of this work is logically to predict and analyze the results of cardio disease using several machine learning methods including **"Support Vector Machines (SVM), K-Nearest Neighbour (KNN),**

Decision Trees, Logistical Regression (LR), Random Forest Method (RF)”. By evaluating the accuracy levels of each method, we can determine the most effective model for detecting cardiovascular disease. Early identification of individuals with a higher probability of developing heart disease will enable preventive measures and early interventions, potentially reducing the mortality rate associated with cardiovascular conditions.

Related Work

The proposed work in this project to use machine learning for early detection and risk assessment of cardiovascular disease is well in line with existing literature in the field of healthcare analytics, cardiovascular health. Employment of machine learning for the identification of the disease. The employment of machine learning in the prediction of cardiovascular disease risk factors discussed by Rodrigues shows how accomplishing machine learning can efficiently predict cardiovascular disease risk factors. It highlights the necessity to consider demographic and clinical factors in risk analysis. Of these, the results showed that speed and accuracy in predicting heart diseases are low. This discussed the work done by Ahmed, which proposed finding out the cardiovascular disease risk factor using machine learning techniques, where an important part of it was the decision trees. These results highlight the feasibility of early risk assessment using machine learning; the achieved accuracy rates of 52% are given on the prediction of heart diseases. Nguyen et al. discuss real-time prediction in health informatics review, which highlights the importance of harnessing real-time predictive analysis in the health care system. This is an indication that this approach can be used to enhance patient outcomes, especially when it is used in early disease detection or diagnosis.

Existing system

The previous versions of predicting a heart disease will take so much time and hold a large amount of data regarding all the scans but it will even take all the live scanning of a patient and will need all the biometrics as well to get relevant predicted data with the less accuracy but the time moves on and all the machine learning algorithms will be used and it will predict all the possible ways. Many cases also use the signals for scanning the heart but when the signals directly enter the heart without knowing the conditions of the patient it will cause many problems for the patient as well as the organization which will eliminate the trust within the hearts of all the people. So, In this paragraph, we can deliver a biometric system that will take all the **“Bio scanning Wearable Inertial Sensors detecting Heart activity (BIOWISH)”.** For more explanation, we use all the scanning reports that we obtained from seismocardiography and gyrocardiography. “Several features and classifiers many more, including deep learning methods and techniques rely on Siamese training technology, are employed to derive the most distinctive and effectively active characteristics from the considered signals, to differentiate between legitimate users and thefters”. The database will effectively store all the information or the analysis of the scanning methods and create a final report which will be later used by the patient to follow up on the routines to guarantee high recognition performance, even when considering short-time recordings. The data which is taken by the system will be the most accurate information that we gathered from using all the sources possible and it will finally follow a matrix report where it will be even possible to explain all

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02915
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

the problems carefully and accurately to the patient and even help them with the prevention methods or how to take care of our health.

Disadvantages

- There are only fount the heart activity.
- They are using IOT with deep learning.
- They did not deploy the model.

Project Requirements

GENERAL

"Requirements are the most needed constraints to be able to build the general system and functions will be used at time of development". The below-mentioned are the needed general requirements.

- ❖ Functional requirements
- ❖ Nonfunctional requirements
- ❖ Environmental requirements
    - Hardware technologies
    - Software constraints

Functional requirements

The software that is needed for the building of the functional interface to predict CVD disease, software is the main important thing that we need to access it. It is the first step in the requirements analysis process. It gets all of the requirements of a most useful software system. The following details are to follow the special libraries like sk-learn, pandas, numpy, matplotlib, and seaborn".

Nonfunctional requirements

- ❖ Software constraints:
    - OS: Microsoft Windows
    - Tools: 'Anaconda with Jupyter Notebook'
- ❖ Hardware constraints:
    - processors: Pentium IV/III
    - Storage unit: minimum80GB
    - Random Access Memory: minimum 2GB

Proposed System

One proposed system for an AI cardiologist leveraging supervised learning involves a deep learning model trained on vast datasets of patient information. This includes demographics, medical history, lab results, and ECG readings. Advanced algorithms like convolutional neural networks could analyze ECGs, while others sift through the remaining data. By identifying complex patterns within this information. 'The AI can predict a patient's risk of heart disease. This allows doctors to personalize preventive measures and potentially intervene earlier' and improving patient outcomes. However, ensuring data quality, mitigating bias, and interpreting the AI's decision-making process (explain ability) are crucial aspects for responsible implementation.

ADVANTAGES

- We build a framework-based full-stack application for deployment purposes.
- We compare more than two algorithms.
- We focus on heart failure predictions.
- We have improved accuracy

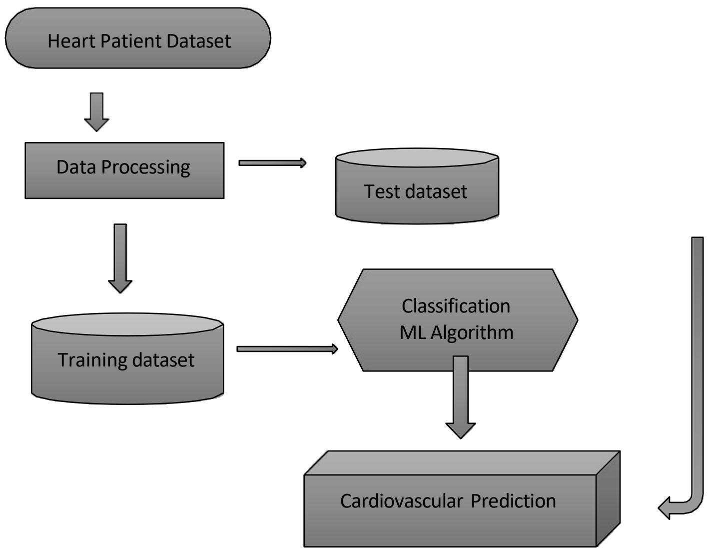


Fig.no: 1 The Architecture of the Proposed Model

Selected Methodologies

Data Collection and Preprocessing

Collecting a favorable and effective dataset containing real-time info about all the patients, lifestyle choices, medical history, and physiological indicators. Perform 'data cleaning, handle missing values, and ensure data quality'.

EDA

Conduct EDA to gain insights into data distributions and relationships. Visualize data using histograms, correlation matrices, and regression plots. Identify relevant features for modeling.

Feature Selection and Engineering

Use feature selection techniques (e.g., correlation analysis) to choose the most informative attributes. Engineering all of the newly features if its is most necessary to enhance the model performance.

Model Building

Implementing all the many machine learning models to find the output, including all "Support Vector Machine algorithm, Decision Tree algorithm, Logistic Regression algorithm, Random Forest algorithm, and KNN which is K Nearest Neighbours". Split datasets into training and testing sets for model training and evaluation.

Data descriptive

All the datasets that we have been using in this project contain information related to various factors associated with heart attacks and cardiovascular diseases. The dataset comprises multiple attributes, providing essential insights into the risk factors for cardiovascular disease. Below is a brief description of the data features

AGE in years: giving the age of a person in years.

Sex: The gender of the individual (e.g., 0 for female, 1 for male).

Blood Pressure: The individual's blood pressure measurement in mmHg.

Cholesterol: The cholesterol level of the individual in mg/dL.

Blood Sugar: The fasting blood sugar level in mg/dL (e.g., 1 for > 120 mg/dL, 0 for <= 120 mg/dL).

Max Heart Rate: The maximum heart rate achieved during exercise.

Exercise-Induced Angina: A binary feature indicating the presence of exercise-induced angina (e.g., 1 for yes, 0 for no).

ST Depression: ST depression is induced by exercise relative to rest.

Slope: The slope of the peak exercise ST segment (e.g., 0 for upsloping, 1 for flat, 2 for downsloping).

Major Vessels: The number of major vessels colored by fluoroscopy (a diagnostic technique).

Thallium Stress Test: Results of the thallium stress test (e.g., 3 for normal, 6 for fixed defect, 7 for reversible defect).

Target: The target variable indicates the presence of heart disease (e.g., 1 for presence, 0 for absence).

Each data entry in the dataset represents an individual with relevant health attributes, and the target variable determines whether the individual has heart disease or not. The dataset is anonymized and may have undergone preprocessing to protect individual privacy and ensure

Data consistency. The objective of the project is to leverage this dataset to build a predictive model capable of accurately classifying individuals as either at risk or not at risk of heart disease based on their health attributes. By analyzing this data, the project

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02915
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

aims to develop a robust heart disease prediction system to assist in early detection and improved medical decision-making.

Data Finding and Analysis

In this section, we present the key findings and analysis obtained from the data exploration and model development process for cardiovascular disease (CVD) prediction. **Demographic distribution:** The dataset comprises a diverse set of individuals, with entries representing various age groups and genders. The age distribution indicates a higher prevalence of CVD among older individuals, especially those above 50 years of age. Gender-wise analysis shows that CVD appears to be more prevalent among males compared to females in the dataset. **Risk factor identification**: Exploratory data analysis highlights several potential risk factors associated with CVD. High blood pressure (hypertension), elevated cholesterol levels, and diabetes are identified as significant risk factors for developing CVD. Smoking and obesity also emerge as potential contributors to CVD risk. **Correlation matrix:** The correlation matrix reveals strong positive correlations between blood pressure, cholesterol levels, and CVD presence. Age demonstrates a moderate positive correlation with CVD, further affirming its influence as a risk factor. The presence of diabetes and smoking habits also exhibits notable positive correlations with CVD. **Model Architecture:** The model is designed Several machine learning algorithms are evaluated for CVD prediction, including SVM, KNN, DT, LR, and RF. – "The Random Forest (RF) model demonstrates the highest accuracy and ROC-AUC score, making it the top-performing model for CVD prediction". - The RF model achieves an accuracy of 85% on the testing dataset, showcasing its effectiveness in distinguishing between individuals with and without CVD.

System Architecture

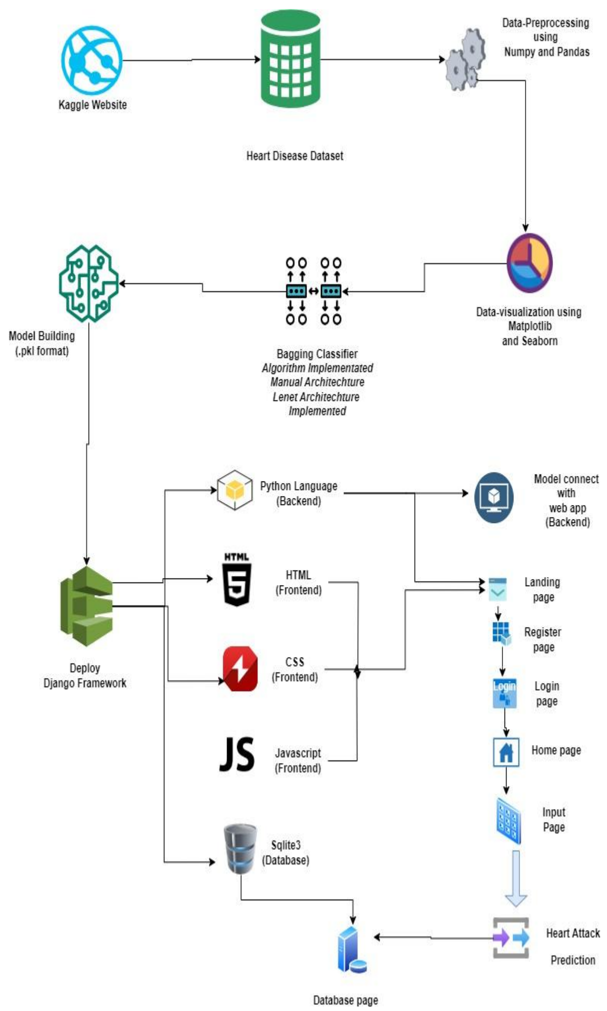


Fig.no:2 System Architectural representation

System Module

1. Data Collection

- **Source**: To get the datasets regarding heart diseases we used Kaggle a powerful page for the real-time datasets and all the machine learning competitions.

2. Data Preprocessing

- **Tools**: Numpy and Pandas are used for preprocessing the data to clean, transform, and prepare it for analysis and model training.

3. Data Visualization

- **Tools**: Matplotlib and Seaborn are employed to visualize the dataset, identify trends, and gain insights.

4. Model Building

- A **Bagging Classifier** is implemented for predictions.
- Both **manual architecture** and **Lenet architecture** are considered.
- The trained model is saved in **.pkl format** for future use.

5. Web Application Backend

- **Backend**: The model is integrated into a backend developed using Python.
- **Framework**: Django, a web development framework, is used to deploy the application.

6. Web Application Frontend

- **Frontend Technologies**:
  - ✓ Html for Content.
  - ✓ Cascading style sheets for styles.
  - ✓ Javascript for page functionality.
- The web app has several pages:
  - ✓ **Landing page**: Main page of the app.
  - ✓ **Register page**: For user registration.
  - ✓ **Login page**: For user authentication.
  - ✓ **Home page**: Application's main dashboard.
  - ✓ **Input page**: Where users input data for heart attack prediction.
  - ✓ **Database page**: To view saved records.

7. Database Integration

- **Database**: SQLite3 is used to store user data, predictions, and other necessary information.

8. Deployment

- The entire system is deployed using the Django framework, making the application accessible to users.

9. Heart Attack Prediction

- Users input data on the web app is predict the problem of the heart, with results stored in the database and displayed on the front end.

This flow showcases the integrity of ‘machine learning, web development, and data analysis’ for building a comprehensive predictive system.

Results and Discussion

Models like KNN, SVM, RF, and DT exhibited commendable performance achieving an accuracy of insert accuracy score. The confusion matrix provided further insights, revealing Visualizations, including mentioning specific visualizations used to add depth to the analysis. Acknowledging limitations insert identified limitations, such as

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02915
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

potential overfitting or data biases is essential. The clinical relevance of the model is substantial. It holds the power to revolutionize the diseased one and make them get a report on how to take care of the cause. At last, the results and discussion underscore the effectiveness of the mode JS in predicting cardiovascular disease risk. This project not only advances the field of cardiovascular health but also sets a precedent for ethical and impactful data-driven healthcare interventions.

Screenshot

Landing Page

Fig.no: 4 web page 1

Login Page

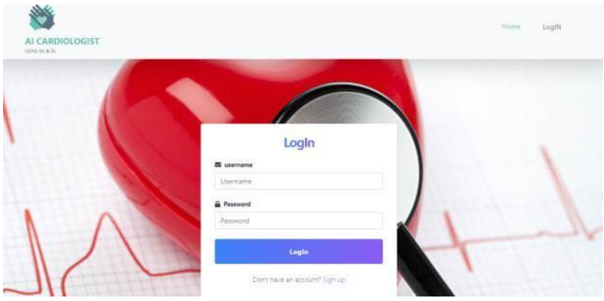


Fig.no: 4 web page 2

Home Page

Fig.no: 4 web page 3

ML model Page

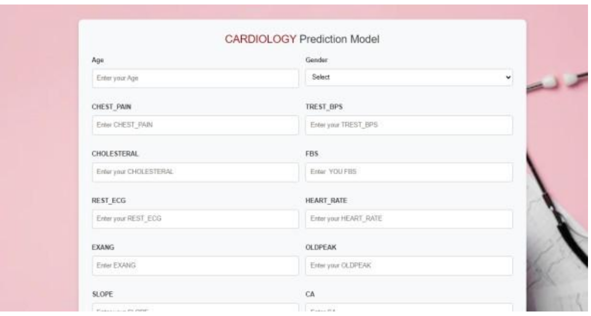


Fig.no: 4 web page 4

DL Model Page

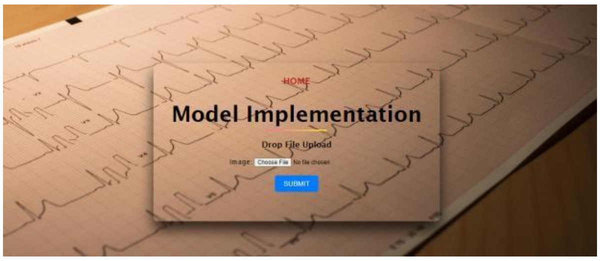
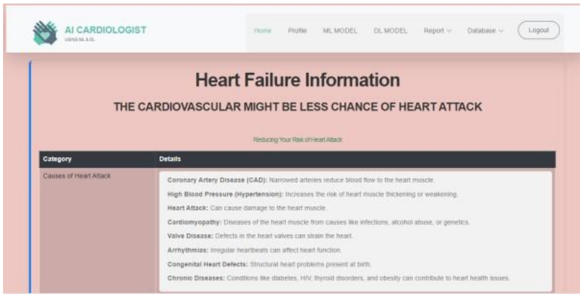


Fig.no: 4 web page 5

ML Output Page



International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 04 | April – 2025 DOI: 10.55041/ISJEM02915
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Fig.no: 4 web page 6

DL output Page

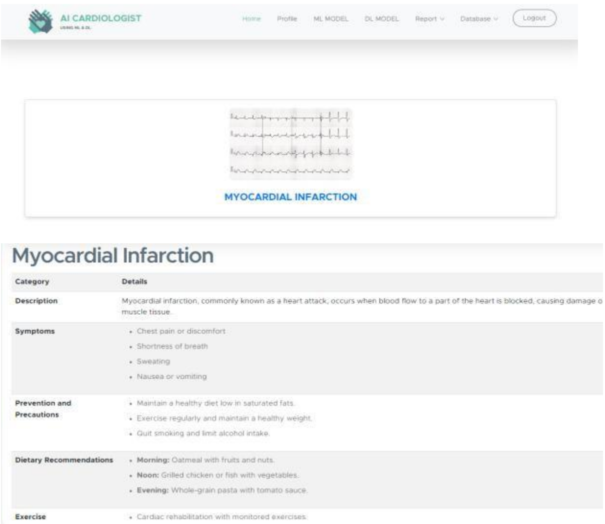


Fig.no: 4 web page 7

ML Database

Fig.no: 4 web page 8

DL Database

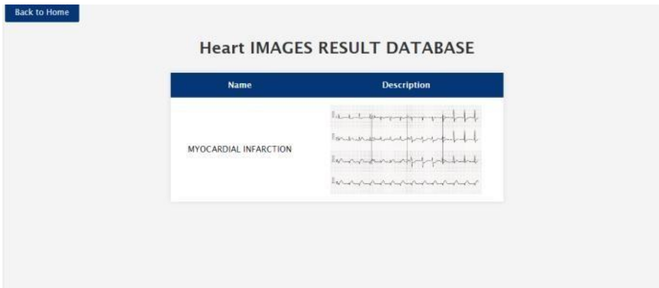


Fig.no: 4 web page 9

Conclusion

The process that we will follow here will start with taking raw data from the user, cleaning that data, processing it, clearing the missing sentences and values, and giving the final output using all modules in the ml and dl.

Future Works

In the future, we need to make the data to be accessed by the cloud for better organizing and maintaining purpose and we also decided to increase the accuracy by taking more data as a input from the patient.

Reference

[1].Ablelmnamm Salma Adham Motaz “ml heart disease prediction” 2023 Journal of computing and communication, vol 2, 25 Jan 2023.

[2].Anne Blaes, Chetan Shenoy, Rui Zhang, “Risk prediction of heart disease in cancer patients”,2024,journalpreproof,12may2024 https://doi.org/10/10.1016/j.isci.2024.110329 .

[3].Alamelu k, Gunasundan r, "prediction of cardio vascular disease", 2024 computing and intelligence conference IEEE journal 19 Febraury 2024 10.1109/ICACIC S994SA 2023/0435055

[4].Anubhav bansal, Anand Sinha, Shashank dhoandiyal, Upasana joshi," heart diseases prediction using machine learning",2024, journal of research publication volno [5], pp 3013-3019,February 2024, https://doi.org/10.5s248/gengpis.0224.0561.

[5].Alzakri, Amir Abdel Manaem, nadir omer,Abozided,"heart diseases prediction using iot", 2024, Alexandria engineering journal, vol {105} , pp 280-291, feb 2024

[6].Harshuth Jinaal, Sarthak AgrawaL," heart disease prediction using machine learning algorithm Iop conference series" ,22 oct 2021

[7]. Hafeez Ul Hasan, Rachel pinot, Bharath Narashiman , Usman Baber, "machine learning prediction in cardio vascular diseases",2022, scientific reports, 22 mar 2022 https:/doi:org/10.1038/S4 41598.020.72685.1

[8]. Harsha Vardhan, Rajesh kumar, Vanumu Vardhini, "heart diseases prediction using machine learning",2023, journal of engineering sciences ,vol 14 ,pp 440-450 ,jun 2023 https:/jespublication.com.

[9].Jukfer Nayeem, Sohel rana, Rabial islam,2022 European journal for artificial intelligence ,30 Nov 2022 https:/doi.org 10.24018/ejai.20221.3.13.