PREDICTION OF CARDIAC DISEASE

A PROJECT REPORT

Submitted by,

LISHA SHREE N - 20201ISI0011
HIMAJA H B - 20201IST0028
PUSHPA LATHA K - 20201IST9004

Under the guidance of,

Dr. R VIGNESH

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

INFORMATION SCIENCE AND TECHNOLOGY

[Internet Technologies]

At

PRESIDENCY UNIVERSITY

BENGALURU

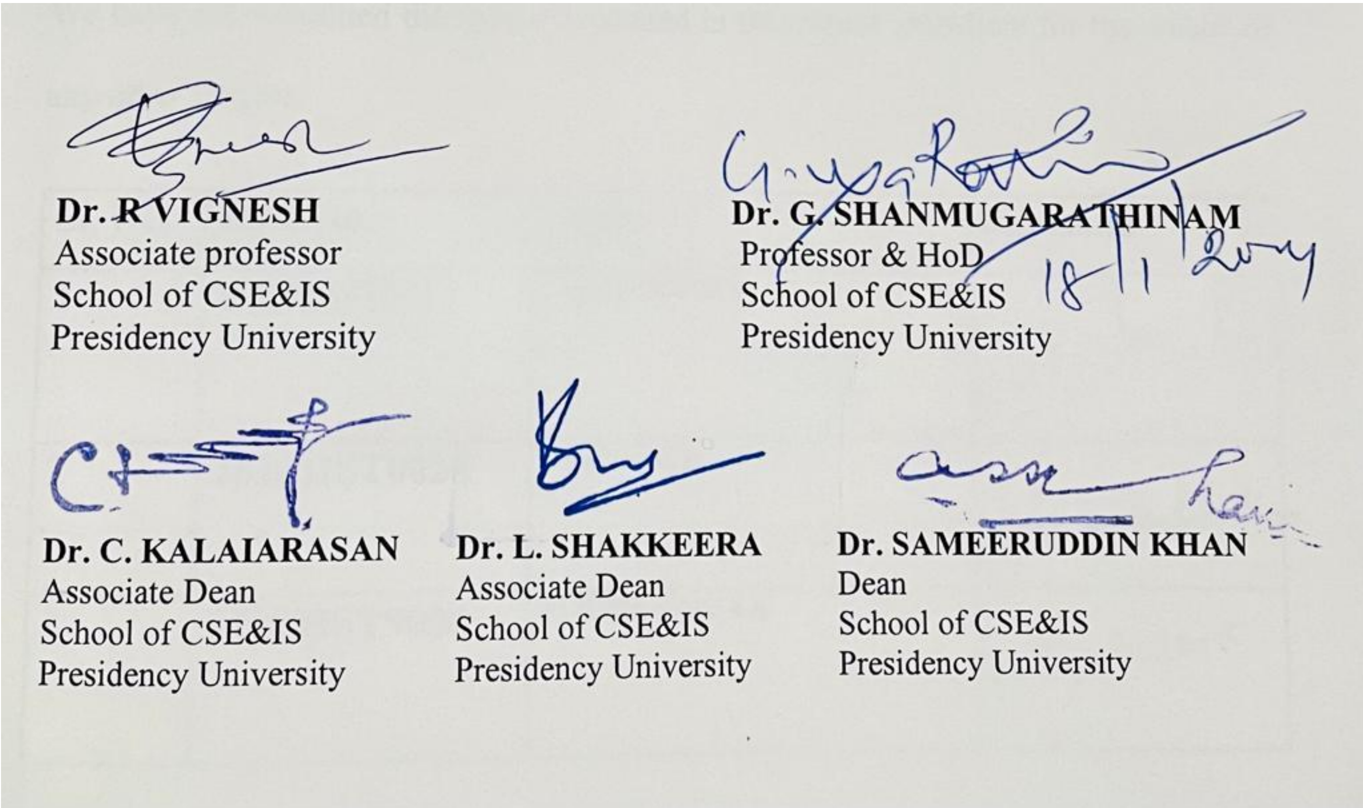
JANUARY 2024

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the Project report **"PREDICTION OF CARDIAC DISEASE"** being submitted by "LISHA SHREE N, HIMAJA H B, PUSHPA LATHA K" bearing roll number(s) "20201ISI0011, 20201IST0028, 20201IST9004" in partial fulfilment of requirement for the award of degree of Bachelor of Technology in Information Science and Technology [Internet Technologies] is a bonafide work carried out under my supervision.



Dr. R VIGNESH
Associate professor
School of CSE&IS
Presidency University

Dr. G. SHANMUGARATHINAM
Professor & HoD
School of CSE&IS
Presidency University
18/1/2024

Dr. C. KALAIARASAN
Associate Dean
School of CSE&IS
Presidency University

Dr. L. SHAKKEERA
Associate Dean
School of CSE&IS
Presidency University

Dr. SAMEERUDDIN KHAN
Dean
School of CSE&IS
Presidency University

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **PREDICTION OF CARDIAC DISEASE** in partial fulfilment for the award of Degree of **Bachelor of Technology** in Information Science and Technology [Internet Technologies], is a record of our own investigations carried under the guidance of **Dr. R VIGNESH, Associate Professor, School of Computer Science & Engineering, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Sl. No.	Roll. No.	Name	Signature
1	**20201ISI0011**	LISHA SHREE N	Lisha Shree .N
2	**20201IST0028**	HIMAJA H B	Himaja. H.B.
3	**20201IST9004**	PUSHPA LATHA K	Pushpalatha K.

ABSTRACT

The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which is able to provide data that aids medical experts in early detection of fatal diseases and therefore, increase the survival rate of patients significantly. Heart disease is the Leading cause of death worldwide. With the rampant increase in the heart stroke rates at juvenile ages, we need to put a system in place to be able to detect the symptoms of a heart stroke at an early stage and thus prevent it. It is impractical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a heart disease. So we proposed a system with the help of machine learning techniques and algorithms like Logistic Regression, KNN, SVC, Random Forest, Decision Tree, XGB Classifier and Naïve Bayes to predict Heart Disease based on different parameters entered by the user in the front end

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Dean, School of Computer Science & Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We record our heartfelt gratitude to our beloved Associate Deans **Dr. Kalaiarasan C and Dr. Shakkeera L,** School of Computer Science & Engineering & Information Science, Presidency University and **Dr G. SHANMUGARATHINAM**. Head of the Department, School of Computer Science & Engineering & Information Science, Presidency University for rendering timely help for the successful completion of this project.

We are greatly indebted to our guide **Dr. R VIGNESH, Associate professor** School of Computer Science & Engineering & Information Science, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the University Project-II Coordinators **Dr. Sanjeev P Kaulgud, Dr. Mrutyunjaya MS** and also the department Project Coordinator **Dr. Sampath A K.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

LISHA SHREE N

HIMAJA H B

PUSHPA LATHA K

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER-1
INTRODUCTION

People in the modern world struggle with extreme stress and anxiety as a result of their restless schedules and regular assignments. On top of that, there is growing concern about people who become addicted to long-term habits like smoking cigars or drinking gutka, which can result in a number of chronic illnesses like cancer, heart disease, liver issues, and kidney failure. For well- known physicians, treating and curing these chronic illnesses is a major challenge.

IT specialists have taken notice of this new difficulty and are assisting in the early detection and treatment of such diseases. Every person is different from the nextin terms of appearance, behaviour, and blood pressure and pulse rate readings. Medical professionals typically define a healthy blood pressure range of 120/80to 140/90 mmHg and a healthy pulse rate of 60 to 100 bpm.

The health sector uses a range of machine learning methods and tools in the market today to forecast chronic illnesses. In spite of these efforts, scientists have found certain shortcomings and are looking for more precise predictive algorithms to identify chronic illnesses in people early on and potentially save lives. Thus, based on user-input parameters at the front end, we propose a systemthat uses machine learning techniques and algorithms, such as XGB

Classifier, K-Nearest neighbour , Support Vector Classifier, Random Forest, Decision Tree, and Naive Bayes.

CHAPTER-2

LITERATURE SURVEY

[1] P. Kola Sujatha and K. Mahalakshmi.

Performance evaluation of supervised machine learning algorithms in predictionof heart disease.

2020 IEEE International Conference for Innovation in Technology (INOCON),pages 1– 7, 2020.

Decision Tree, Naive Bayes, Random Forest, Support Vector Machine, K Nearest Neighbour, and logistic regression algorithms are used in the aforementioned paper to detect the existence of heart illness. The algorithms' performance was evaluated by utilizing metrics including F1-score, AUC, accuracy, and precision. According to the experimental results, Random Forest outperforms other supervised machine learning algorithms in the prediction of heart disease, with an accuracy of 83.52%. Random

Forest classifiers have an F1-Score, an AUC, and an accuracy score of 84.21%, 88.24% and 88.89%, respectively.

[2] **Pahulpreet Singh Kohli and Shriya Arora.** Application of machinelearning in disease prediction.

2018 4th International Conference on Computing Communication and Automation(ICCCA), pages 1–4, 2018.

In this work, we use three distinct disease databases (heart, breast cancer, and diabetes) from the UCI repository to forecast diseases using various classificationtechniques, each with a unique set of benefits. Backward modeling with the p
value test was used to choose features for each dataset. The study's findings support the notion that machine learning can be used to diagnose diseases early on.

[3] Abderrahmane Ed-Daoudy and Khalil Maalmi.

Real-time machine learning for early detection of heart disease using big data approach. pages 1–5, 04 2019.

This work proposes a real-time heart disease prediction system using Apache Spark, a powerful large-scale distributed computing platform that can be effectively used for in-memory computations to stream data events against machine learning. The two primary sub-components of the system are streaming processing and data storage andvisualization. In the first, a classification model is applied to data events using Spark MLlib and Spark streaming to forecast heart illness. The massive amount of created data is stored by the seconds using Apache Cassandra.

[4] A. Lakshmanarao, A. Srisaila, and Srinivasa Tummala.

Heart disease prediction using feature selection and ensemble learning techniques. pages 994–998, 02 2021.

They presented a brand-new machine learning model for predicting cardiac disease in this research. Two distinct datasets from Kaggle and UCI were used to evaluate the suggested approach. To determine the best features, we used feature selection approaches after applying sampling techniques to the unbalanced dataset. Subsequently, an ensemble classifier was used to apply many classifier models and achieve good accuracy. Experiments conducted on two datasets have demonstrated the effectiveness of the suggested model in predicting cardiac disease.The implementations were entirely done in Python.

[5] Alperen Erdoğan and Selda Guney.

Heart disease prediction by using machine learning algorithms.
2020 28th Signal Processing and Communications Applications Conference (SIU),pages 1–4, 2020.

These days, heart disease is one of the major illnesses that kills the majority of patients. Heart disease is extremely difficult to diagnose medically. Even though cardiac conditions are medically diagnosed, they might be mistaken for other conditions that present with similar symptoms, such as nausea, palpitations, chest pain, and shortness of breath. This complicates medical diagnosis of heart problems.In this work, machine learning algorithms were used to identify the presence of heart disorders. The patient data included in this study were ranked in accordance with how they affected the success rate. This paper suggests a way for figuring out the weight coefficient. Based on the findings of the suggested strategy, 86,90% of the patients' features were successfully collected.

[6] Shaik Farzana and Duggineni Veeraiah.

Dynamic heart disease prediction using multi-machine learning techniques.

2020 5th International Conference on Computing, Communication and Security(ICCCS), pages 1–5, 2020.

The health care industry generates a lot of data, therefore in order to process it and produce useful results, we need to apply modern techniques. These approaches also help us make informed decisions based on the data and produce the right results. The primary issue and one of the main reasons for the number ofdeaths worldwide is heart disease. In this paper, machine learning methods including Gaussian Naive Bayes, Random Forest, K-Nearest Neighbor, and SVC are used to create an efficient framework for heart disease prediction. Thirteen features are used by the framework, including cp, age, gender, blood pressure, cholesterol, and obesity. We are using a user-friendly approach that is divided into several stages. First, we choose the algorithm to run on the dataset file and uploadit. After that, an algorithm is used to pick the accuracy along with a graph, and thedataset is trained to create a modal for the one with the highest frequency

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

In spite of tremendous advancements, there are still several aspects of Machine learning based cardiac disease prediction that require research and development.

To create more accurate and dependable prediction models, it is essential to comprehend these gaps. The following are some unmet needs and current approaches in the field of heart disease prediction:

3.1. Limited Variability within the Sets:

Research Gap: There is a dearth of diversity in terms of demographics, lifestyle, and geographic representation in many of the datasets currently usedfor the prediction of heart disease. Current techniques: Scholars have made use of well-known datasets, including those from the Cleveland Heart Disease and the Framingham Heart Study. The model has to be improved with more varied datasets from various populations in order to become more broadly applicable.

3.2. Choosing Features and Their Significance:

1. Research Gap: It's still challenging to identify the key characteristics thatbest predict heart disease. Understanding the characteristics that significantly enhance accurate forecasting is essential for clinical applicability as well as model interpretability.

2. Current Approaches: Recursive Feature Elimination (RFE), feature importance derived from tree-based models, and domain knowledge driven feature selection are a few of the feature selection strategies that have been used. But more reliable and automated techniques are required.

3. Research Gap: Biased models can result from imbalanced datasets, where one class (such as the presence of heart disease) is noticeably underrepresented.

4. Current Methods: To address class imbalance, methods such as undersampling, oversampling, and the use of synthetic data (SMOTE) have been used. Nevertheless, more research is required to find the best strategy for imbalanced datasets related to heart disease.

Unbalanced Collections:

1. Research Gap: Biased models can result from imbalanced datasets, where one class (such as the presence of heart disease) is noticeably underrepresented.

2. Current Approaches: To address class imbalance, methods such as under sampling, oversampling, and the use of synthetic data (SMOTE) have beenused.

Model Interpretability:

1. Research Gap: Interpretability issues prevent many machine learning models—especially complex ones like ensemble methods—from being widely used in clinical settings.
2. Current Approaches: Interpretability strategies have been including model-agnostic approaches, LIME, and SHAP values.

Temporal Elements:

1. Research Gap: The onset of heart disease is a dynamic process that is impacted by time. Current models frequently ignore the data's temporal dimensions.

CHAPTER-4

PROPOSED MOTHODOLOGY

- A Heart Disease Dataset is extracted.
- The dataset is taken and pre-processed with many Machine Learningtechniques.
- The pre-processed data is again divided into testing data.
- The prediction model is built using machine learning algorithms likeLogistic Regression, K Nearest Neighbour, Support
- VectorClassifier, Random Forest ,Decision Tree, Naive Bayes and XGB
- The model is trained by using training dataset.
- Once the model is been trained completely it should be tested toget accurate result.
- After getting accurate result the accuracy is calculated.
- Among the above algorithms which provides most accurate results isfurther taken for finalising the model.
- The model is converted into pickle model (binary data) and then finalised.
- Model is developed using Front End with the use of Flask andHTML method.
- Now user will enter all the required parameters to predict the cardiacdisease in the front end.
- The details given by the user is taken as input (front end) to our final algorithm to estimate whether the person has the cardiac disease or does nothave a disease.
- And, the estimated output is displayed, shown on the front end.

CHAPTER-5

OBJECTIVES

Addressing critical gaps in machine learning-based heart disease prediction is imperative for improving model reliability and accuracy. Notably, currentdatasets lack diversity in demographics and lifestyle, emphasizing the need for more varied datasets to enhance model generalizability. Feature selection remains challenging, and while methods like Recursive Feature Elimination are utilized, there is a demand for more reliable automated techniques. Dealing with imbalanced datasets, a common issue, requires further investigation into the most effective strategies beyond current methods.Model interpretability is another gap, with a need for a better balance between complexity and interpretability, especially for complex models like ensembles. Additionally, incorporating the temporal dimension in models is essential, recognizing heart disease as a dynamic process impacted by time. In order to accurately reflect the dynamic nature of the disease, efforts shouldbe directed on creating models that take into account how risk variables change over time.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

SYSTEM IMPLEMENTATION

Data Flow Diagram

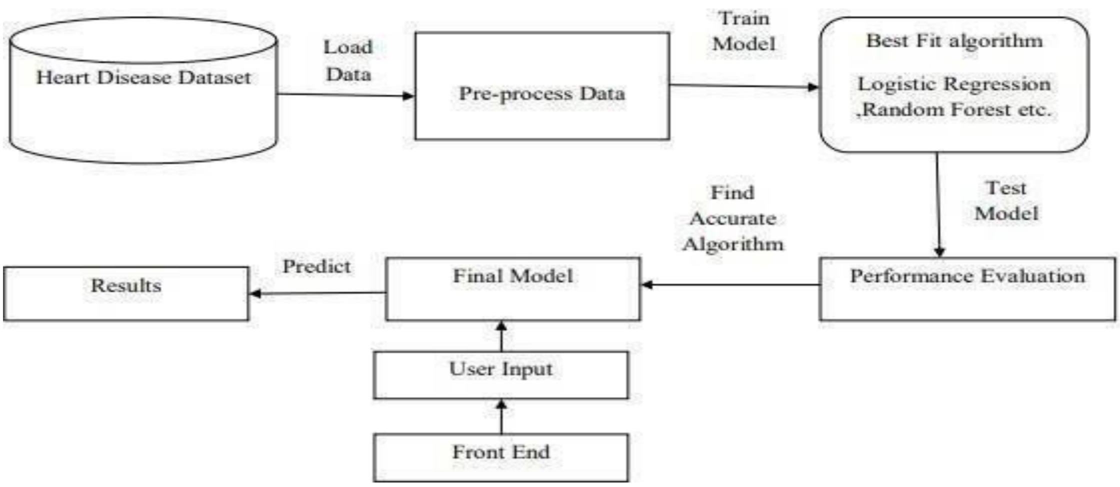


Fig 6.1. Data Flow Diagram

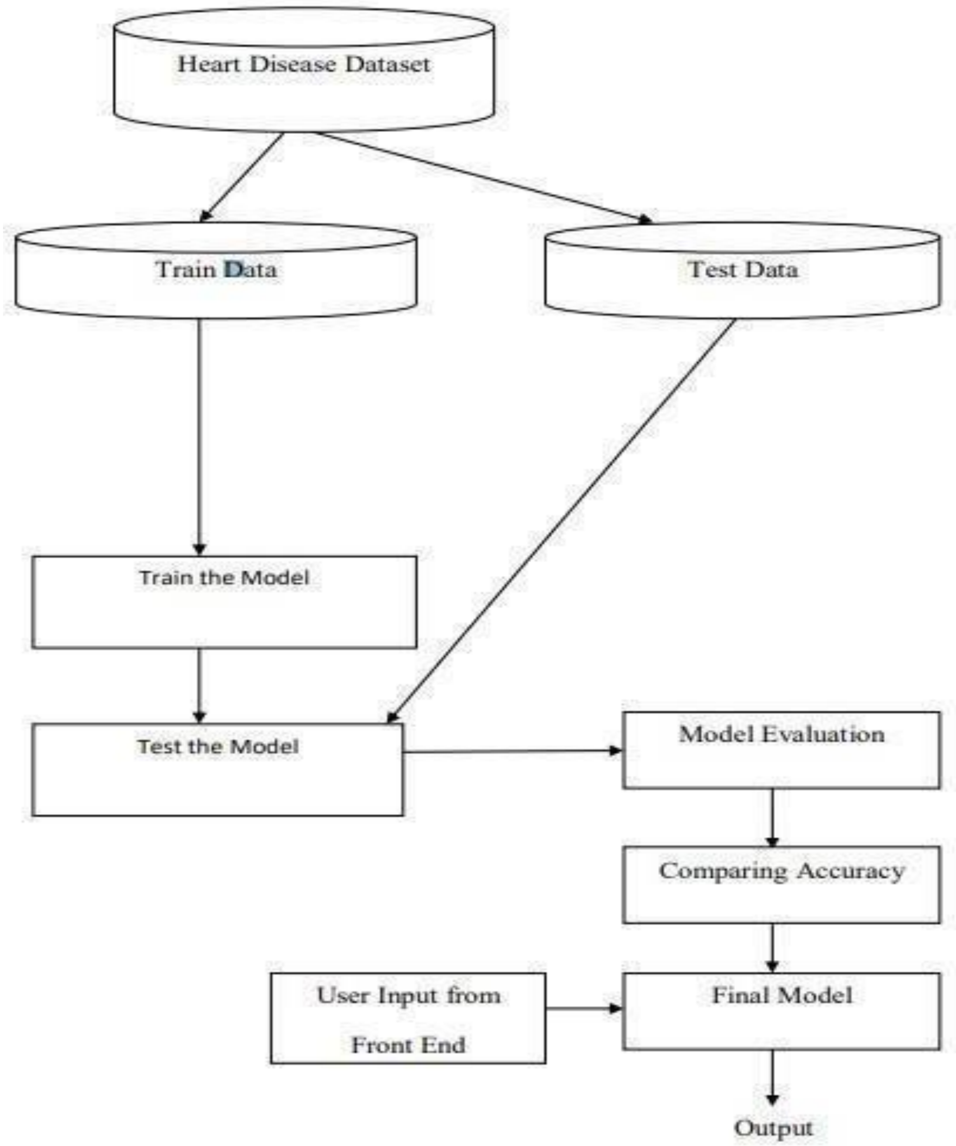


Fig.6.2.System architecture

The model evaluation procedure is a crucial phase in the heart disease prediction system's process to guarantee its correctness and dependability. First, a broad dataset covering lifestyle, health history, and demographics is obtained. To aidin the training and assessment of the model, this dataset is then divided into subsets for testing and training. During the training phase, the model is trainedon the training dataset using techniques such as Random Forest, with parameters being adjusted for best results. The model's capacity to generalize to new, unseen data is then evaluated by testing it on the reserved dataset.

Metrics including accuracy, precision, recall, and F1 score are computed duringthe assessment to measure the model's effectiveness. The model's ability to forecast cardiac disease can be evaluated by comparing these metrics. Although accuracy is a crucial indicator that indicates how accurate predictions are generally, other metrics offer a more detailed picture, particularly when takingany class imbalances in the dataset into account. .

In addition to quantitative measurements, the model evaluation phase produces qualitative insights into the model's advantages and any shortcomings. These observations are essential for improving the model's accuracy and predictive power. This technique is iterative, thus ongoing improvement is guaranteed based on input from the testing stage.

In this entire process, the user interface (UI) is essential since it acts as a safe portal for consumers to enter their data. In addition to efficiently gathering userinput, the user interface (UI) should provide the model's predictions in an easilycomprehensible manner. In order to improve the system's usability, resolve anydisparities, and further refine the model, user feedback from the user interface is crucial. The system's continuous optimization and user-needs adaption are guaranteed by this closed feedback loop between users and the system.

Use Case Diagram

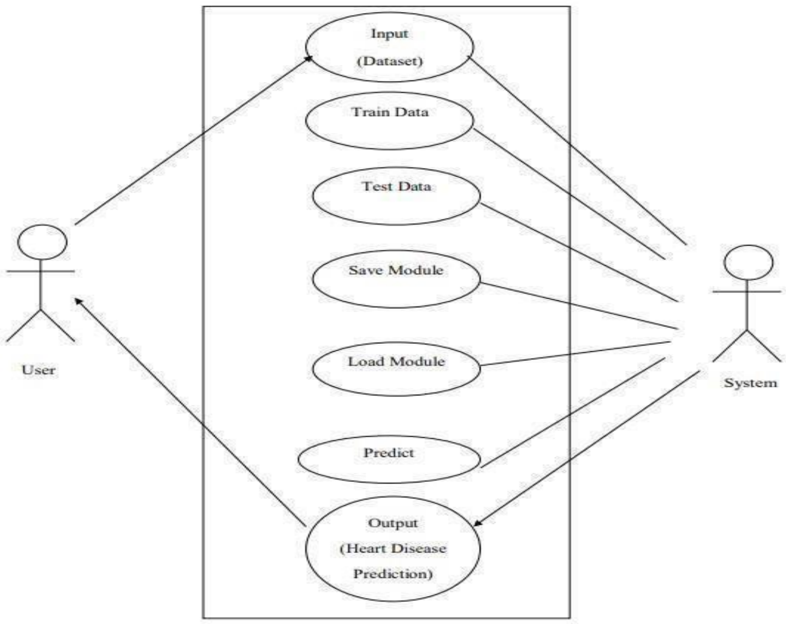


Fig 6.3. Use Case Diagram

The heart disease prediction system begins by collecting relevant data fromthe

user. Once the data is gathered, the system initiates an analysis to extract meaningful insights. Subsequently, the collected data undergoes preprocessing phase where missing values are addressed, and any necessary adjustments are made to enhance the overall quality of the information. Thesystem then proceeds to build a predictive model using Machine Learning algorithms, incorporating techniques like Random Forest Classifier, Decision Tree Classifier(tree-shaped model), KNN(K-Nearest Neighbors), Support Vector Classifier (Support Vector Machines in general), Logistic Regression, Naive Bayes, and XGB Classifier.

Upon constructing the model, a critical step involves evaluating and testingits accuracy. This ensures that the predictive capabilities of the model are reliable & effective. The user actively participates in the decision-making process by finalizing the model with the best accuracy among the evaluated algorithms. This user-driven selection is crucial for tailoring the system specific preferences or requirements.

The user can feed fresh data into the system when the model is complete. After that, the predictive model analyze the data and projects the probability of heartdisease using the well-established algorithms. The user is then presented with the results, which provide insightful information about the person's potential danger. To put it simply, the system offers a user-friendly and precise tool for heart disease prediction by seamlessly integrating advanced analysis, machine learning capabilities, and human input.

Activity Diagram

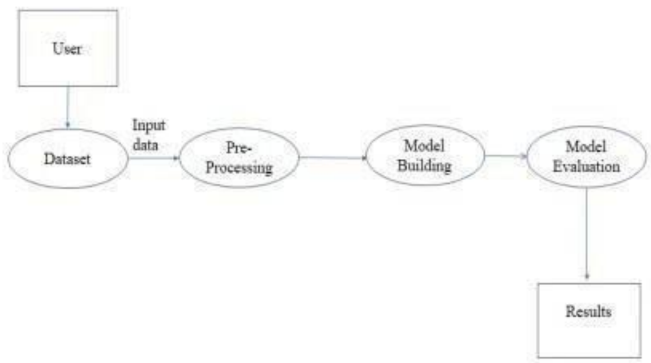


Fig 6.4. Activity Diagram

In the workflow of the heart disease prediction system, users initiate the process by providing a dataset to the system. This dataset is then subjected to a preprocessing stage aimed at enhancing the overall accuracy of the subsequent predictive model. The preprocessing step involves handling missing values, addressing outliers, and making necessary adjustments to ensure the dataset is optimized for model building.

The system uses a number of algorithms to build the heart disease prediction model after the dataset has been suitably pre-processed. These algorithms, which may include Random Forest Classifier, Decision Tree Classifier(tree-shaped model), KNN(K-Nearest Neighbors), Support Vector Classifier (Support Vector Machines in general), Logistic Regression, Naive Bayes, and XGB Classifier, contribute to the creation of a robust and versatile predictive model.

After model construction, an evaluation phase ensues to assess the performanceof each algorithm. The system determines the algorithm that yields the highestaccuracy, and this particular model is then finalized for use. The finalization step involves selecting the model with the best accuracy, ensuring that it is wellsuited for the specific dataset and user requirements.

The user can then utilize the completed model to forecast outcomes using either fresh or preexisting data. The model's predictive abilities offer important insights into the probability of heart disease, making it a useful and effective tool for risk assessment. In conclusion, this user-centric methodology—which includes the provision of datasets, preprocessing, model construction,evaluation, and finalization—leads to the creation of a system that provides users with precise and knowledgeable forecasts regarding heart disease.

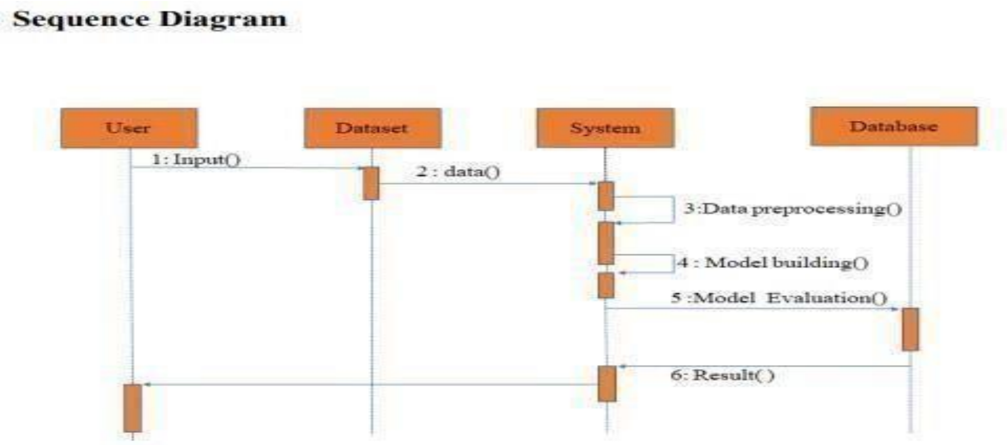


Fig 6.5. Sequence Diagram

The heart disease prediction system operates in a user-friendly manner, beginning with users providing datasets as inputs. The system then efficiently stores the user-provided dataset in its database, ready for subsequent processing. In the preprocessing phase, the stored data undergoes careful handling, addressing missing values and optimizing the dataset for examination in order to improve the following model's accuracy. Following preprocessing, system embarks on the crucial task of model construction. Utilizing various machine learning algorithms, including but not limited to Random Forest Classifier, Decision Tree Classifier(tree-shaped model), KNN(K-Nearest Neighbors), Support Vector Classifier (Support Vector Machines in general), Logistic Regression, Naive Bayes, and XGB Classifier, the system builds a predictive model. This model is subsequently trained using the pre-processed data, enabling it to learn and adapt to patterns within the dataset. Once the model is constructed and trained, an evaluation process ensues to gauge the performance of each algorithm. The system carefully assesses accuracy metrics, ultimately selecting the algorithm that demonstrates the highest accuracy for finalization. This chosen algorithm becomes the basis for the finalized model, tailored to provide optimal predictive capabilities.

Using the completed model to forecast outcomes based on fresh or user supplieddata is the last stage. Users can rely on the model's accuracy to obtain insightfulinformation about the risk of heart disease. In essence, this streamlined process, from dataset input to model finalization, underscores the system's efficacy in providing users with accurate and reliable predictions, contributing to informed decision-making regarding heart disease risk assessment.

Implementation:

In the implementation phase, Python serves as the primary programming language for developing the heart disease prediction system. Python is chosen for its interpreted nature, object-oriented paradigm, and high-level features withdynamic semantics. The language is perfect for scripting and Rapid Application Development because of its built-in data structures, dynamic typing, and dynamic binding. Python's ease of use, readability, and large standard library make it a popular choice, particularly for machine learning.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
52	1	0	125	212	0	1	168	0	1	2	2	3	0
53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
61	1	0	148	203	0	1	161	0	0	2	1	3	0
62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
58	0	0	100	248	0	0	122	0	1	1	0	2	1
58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
54	1	0	122	286	0	0	116	1	3.2	1	2	2	0
71	0	0	112	149	0	1	125	0	1.6	1	0	2	1
43	0	0	132	341	1	0	136	1	3	1	0	3	0
34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
51	1	0	140	298	0	1	122	1	4.2	1	3	3	0
52	1	0	128	204	1	1	156	1	1	1	0	0	0
34	0	1	118	210	0	1	192	0	0.7	2	0	2	1
51	0	2	140	308	0	0	142	0	1.5	2	1	2	1
54	1	0	124	266	0	0	109	1	2.2	1	1	3	0
50	0	1	120	244	0	1	162	0	1.1	2	0	2	1
58	1	2	140	211	1	0	165	0	0	2	0	2	1
60	1	2	140	185	0	0	155	0	3	1	0	2	0
67	0	0	106	223	0	1	142	0	0.3	2	2	2	1
45	1	0	104	208	0	0	148	1	3	1	0	2	1
63	0	2	135	252	0	0	172	0	0	2	0	2	1
42	0	2	120	209	0	1	173	0	0	1	0	2	1
61	0	0	145	307	0	0	146	1	1	1	0	3	0

Fig 6.6. Training Data

The Python IDLE, or Integrated Development Environment, serves as the development environment within the Python ecosystem. IDLE is a software development toolkit that combines multiple tools designed for productive work. These tools include source control capabilities, build, execution, and debugging tools, and a code editor with features like syntax highlighting and auto- completion. A Python shell window for interactive interpretation, auto-completion, syntax highlighting, intelligent indentation, and a rudimentary integrated debugger are among the features of Python IDLE. Python 3.68 is the version that is specially used in this implementation.

HTML (Hypertext Markup Language) is incorporated to structure the web-based interface of the heart disease prediction system. HTML serves as the code defining the structure of the webpage and its content. The language utilizes elements enclosed in opening and closing tags to structure content, such as paragraphs, lists, images, and data tables. An opening tag, a closing tag, and the content inside make up each HTML element. Extra details about the element are contained in attributes, which are used to provide the content more features.The class attribute in HTML is used to provide non-unique identifiers so that style information can be applied.

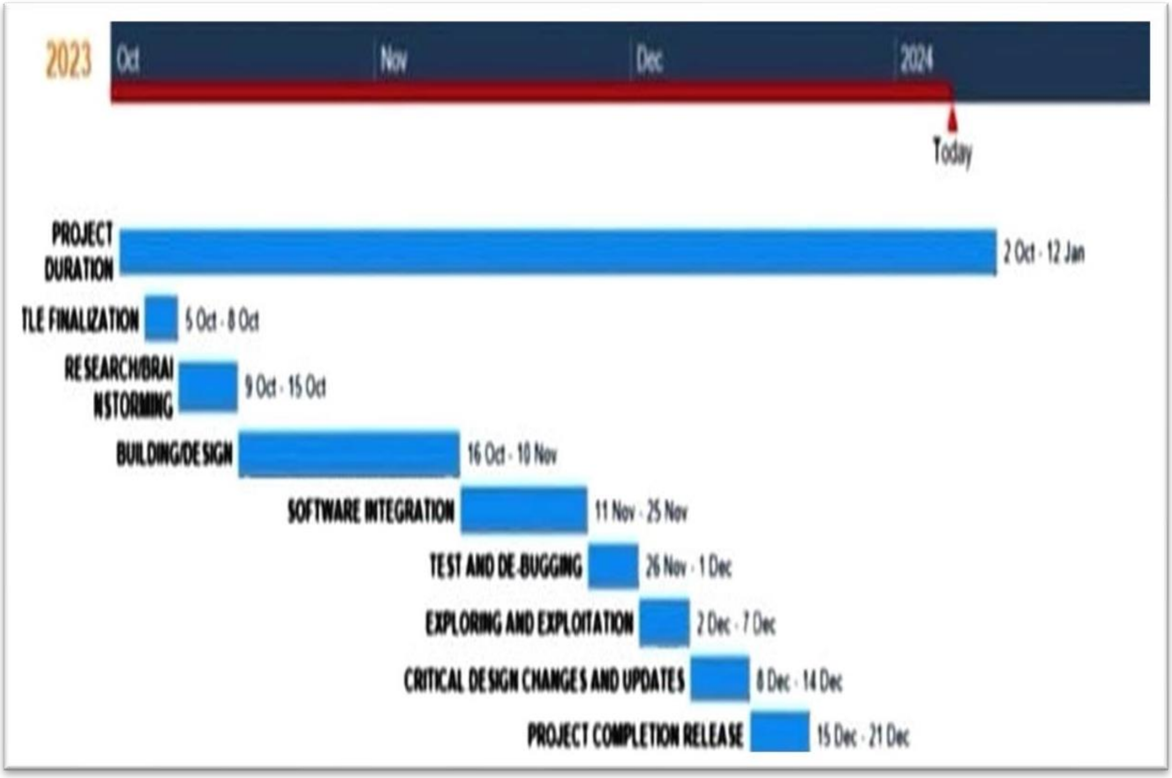
An HTML element's content, closing tag, and opening tag define its anatomy. The content of an element is the text or data that is contained within it; the opening and closing tags indicate the start and finish of the element, respectively. Features that aren't present in the content itself, like class, are added to give more details about the element. There should be a gap between the attribute and the element in proper attribute syntax.In summary, the heart disease prediction system implementation combines the power of Python for Machine Learning algorithms and also data processing along with HTML for creating an user-friendly web interface, demonstrating the synergy between programming languages and web technologies in developing practical applications.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)

PREDICTION OF CARDIAC DISEASE FIG. 7.1



CHAPTER-8

OUTCOMES

The implementation of the heart disease prediction system has yielded several significant outcomes. Foremost among these is the successful creation of an accurate prediction model, achieved through the application of various Machine Learning algorithms, including Random Forest Classifier, Decision Classifier(tree-shaped Tree model), KNN(K-Nearest Neighbors), Classifier (Support Vector Machines
in general), LogisticSupportVectorRegression, Naive Bayes, and XGB Classifier. This ensures a comprehensive and robustmodel capable of providing reliable predictions for heart disease.

Additionally, the system boasts a user-friendly interface, thanks to the incorporation of HTML. This enables users to seamlessly interact with the system, providing datasets and visualizing results with ease. The interface enhances the overall user experience, making the system accessible and intuitive.

Additionally, the method makes use of Python's powerful data processing features. Python's dynamic semantics and high-level built-in data structures areinvaluable for efficiently managing and analyzing the datasets. This effectiveness enhances the overall functionality and dependability of the system. In conclusion, the outcomes of the heart disease prediction system implementation include the successful development of an accurate model, a userfriendly interface, and efficient data processing capabilities. These achievements collectively contribute to a system that is not only reliable and accurate but also accessible and user-centric in its design and functionality.

CHAPTER-9

RESULTS AND DISCUSSIONS

The endeavor to predict and detect heart disease has long been a formidable challenge for healthcare practitioners. With expensive therapies and operations being the norm for treating heart diseases, the importance of early detection cannot be overstated. The ability to predict heart disease in its early stages holdsimmense potential for individuals worldwide, allowing them to take proactive measures and prevent the progression of the condition to a more severe state.

We developed and implemented a system utilizing machine learning techniques and algorithms in response to this pressing healthcare issue. Based on a number of parameters that the user enters at the front end, the system is intended to forecast heart disease. Logistic Regression, KNN (K-Nearest Neighbors), SVC (Support Vector Classifier), Random Forest, Decision Tree, XGB Classifier, and Naive Bayes are among the selected machine learning methods. Every algorithm adds distinct skills to the prediction model, guaranteeing a thorough examination of the input parameters.

Upon implementation and testing, our project demonstrated promising results in the prediction of cardiac disease. The system achieved , commendable accuracy rate of 91.80%, as determined by the Random Forest Classifier. This high level of accuracy is indicative of the effectiveness of the machine learning model in discerning patterns and associations within the dataset, leading to reliable predictions.

In conclusion, the successful implementation of our heart disease prediction system marks a significant step toward addressing the challenges in healthcarerelated to heart diseases. The achieved accuracy of 91.80%, shows how machine learning may help with early detection and prediction, especially when used with the Random Forest Classifier. This system holds promise for empoweringindividuals to make informed decisions about their health and wellbeing, paving the way for proactive measures and improved healthcare outcomes. As technology continues to advance, such predictive models have the potential to revolutionize preventive healthcare practices and contribute to a healthierglobal population.

Sl.No	Models	Accuracy
1	Naïve Bayes	90.16%
2	Logistic Regression	85.25%
3	Random Forest	91.80%
4	KNN	65.57%
5	Decision Tree	78.69%
6	SVC	86.89%
7	XGB Classifier	86.89%

Fig 9.1. Comparison of other algorithms with Random forest

CHAPTER-10
CONCLUSION

In summary, our heart-cardiac prediction system uses a variety of Machine learning techniques to achieve an amazing 91.80% accuracy, with Random Forest being a major contributor. The multi-algorithmic approach provides a comprehensive analysis of user-entered parameters. The userfriendly front end, developed with Flask, HTML, and pymysql, facilitates easy interaction, empowering users to proactively engage with their health data. This system holds immense potential for revolutionizing early diagnosis and intervention in heart disease cases, exemplifying the symbiosis of healthcare and machine learning. Ongoing refinement and exploration of additional features promise to further enhance the system's predictive capabilities, shaping the future ofpreventive healthcare practices globally.

REFERENCES

1. P. Kola Sujatha and K. Mahalakshmi. Performance evaluation of supervised machine learning algorithms in prediction of heart disease. *2020 IEEE International Conference for Innovation in Technology (INOCON)*, pages 1– 7, 2020.
2. Pahulpreet Singh Kohli and Shriya Arora. Application of machine learning in disease prediction. *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pages 1–4, 2018.
3. Abderrahmane Ed-Daoudy and Khalil Maalmi. Real-time machine learning for early detection of heart disease using big data approach. pages 1–5, 04 2019.
4. Lakshmanarao, A. Srisaila, and Srinivasa Tummala. Heart disease prediction using feature selection and ensemble learning techniques. pages 994–998, 02 2021.
5. Alperen Erdoğan and Selda Guney. Heart disease prediction by using machine learning algorithms. *2020 28th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2020.
6. Shaik Farzana and Duggineni Veeraiah. Dynamic heart disease prediction using multi-machine learning techniques. *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, pages 1–5, 2020.

APPENDIX-A
PSUEDOCODE

```
import numpy as np
import pandas as pd


import pymysql
pymysql.install_as_MySQLdb()
import MySQLdb


gmail_list=[]
password_list=[]
gmail_list1=[]
password_list1=[]


from flask import Flask, request, jsonify, render_template


import joblib

# Load the model from the file
model = joblib.load('heart_disease.pkl')


# Use the loaded model to make predictions


app = Flask(_name_)
```

```
@app.route('/')
def home():
    return render_template('register.html')


@app.route('/register',methods=['POST'])
def register():


    int_features2 = [str(x) for x in request.form.values()]

    r1=int_features2[0]
    print(r1)

    r2=int_features2[1]
    print(r2)
    logu1=int_features2[0]
    passw1=int_features2[1]




   # if int_features2[0]==12345 and int_features2[1]==12345:

    import MySQLdb


# Open database connection
    db = MySQLdb.connect(host="localhost",user="root",password='',database="ddbb" )

# prepare a cursor object using cursor() method
    cursor = db.cursor()
    cursor.execute("SELECT user FROM user_register")
```

```
    result1=cursor.fetchall()



    for row1 in result1:
                print(row1)
                print(row1[0])
                gmail_list1.append(str(row1[0]))




    print(gmail_list1)
    if logu1 in gmail_list1:
        return render_template('register.html',text="This Username is Already in Use ")

    else:





# Prepare SQL query to INSERT a record into the database.
            sql = "INSERT INTO user_register(user,password) VALUES (%s,%s)"
            val = (r1, r2)

            try:
  # Execute the SQL command
                        cursor.execute(sql,val)
  # Commit your changes in the database
                        db.commit()
            except:
   # Rollback in case there is any error
                        db.rollback()

# disconnect from server
            db.close()
            return render_template('register.html',text="Succesfully Registered")
@app.route('/login')
```

```
def login():
    return render_template('login.html')




@app.route('/logedin',methods=['POST'])
def logedin():

    int_features3 = [str(x) for x in request.form.values()]
    print(int_features3)
    logu=int_features3[0]
    passw=int_features3[1]



    import MySQLdb



# Open database connection
    db = MySQLdb.connect(host="localhost",user="root",password='',database="ddbb")

# prepare a cursor object using cursor() method
    cursor = db.cursor()
    cursor.execute("SELECT user FROM user_register")
    result1=cursor.fetchall()

    for row1 in result1:
                print(row1)
                print(row1[0])
                gmail_list.append(str(row1[0]))




    print(gmail_list)


    cursor1= db.cursor()
    cursor1.execute("SELECT password FROM user_register")
```

```
    result2=cursor1.fetchall()

    for row2 in result2:
                print(row2)
                print(row2[0])
                password_list.append(str(row2[0]))



    print(password_list)
    print(gmail_list.index(logu))
    print(password_list.index(passw))

    if gmail_list.index(logu)==password_list.index(passw):
        return render_template('index.html')
    else:
        return render_template('login.html',text='Use Proper Username and Password')
@app.route('/production')
def production():
    return render_template('index.html')



@app.route('/production/predict',methods=['POST'])
def predict():
    '''
    For rendering results on HTML GUI
    '''
    int_features = [str(x) for x in request.form.values()]
    a=int_features








age = int(a[0])
```

```
    sex=int(a[1])
    cp = int(a[2])
    trestbps =int(a[3])
    chol = int(a[4])
    fbs=int(a[5])
    restecg=int(a[6])
    thalach=int(a[7])
    exang=int(a[8])
    oldpeak=float(a[9])
    slope=int(a[10])
    ca=int(a[11])
    thal=int(a[12])

    df=[age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal]


# Print the output.
    print(df)
    df=np.array(df).reshape(1,-1)

    prediction=model.predict(df)
    prediction=int(prediction)
    print(prediction)



if(prediction==1):
        return render_template('index.html',prediction_text='The Person has the Heart Disease')
    else:
       return render_template('index.html',prediction_text='The Person not have the Heart Disease')


  if _name_ == "_main_":
    app.run(debug=False)
```

backendcode

```
# for numerical computing
import numpy as np


# for dataframes
import pandas as pd


#for plotting
import matplotlib.pyplot as plt
import seaborn as sns

# Ignore Warnings
import warnings
warnings.filterwarnings("ignore")

# to split train and test set
from sklearn.model_selection import train_test_split

# Machine Learning Models
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from xgboost import XGBClassifier


from sklearn.metrics import accuracy_score

# to save the final model on disk

data=pd.read_csv('heart.csv')
print(data.shape)
```

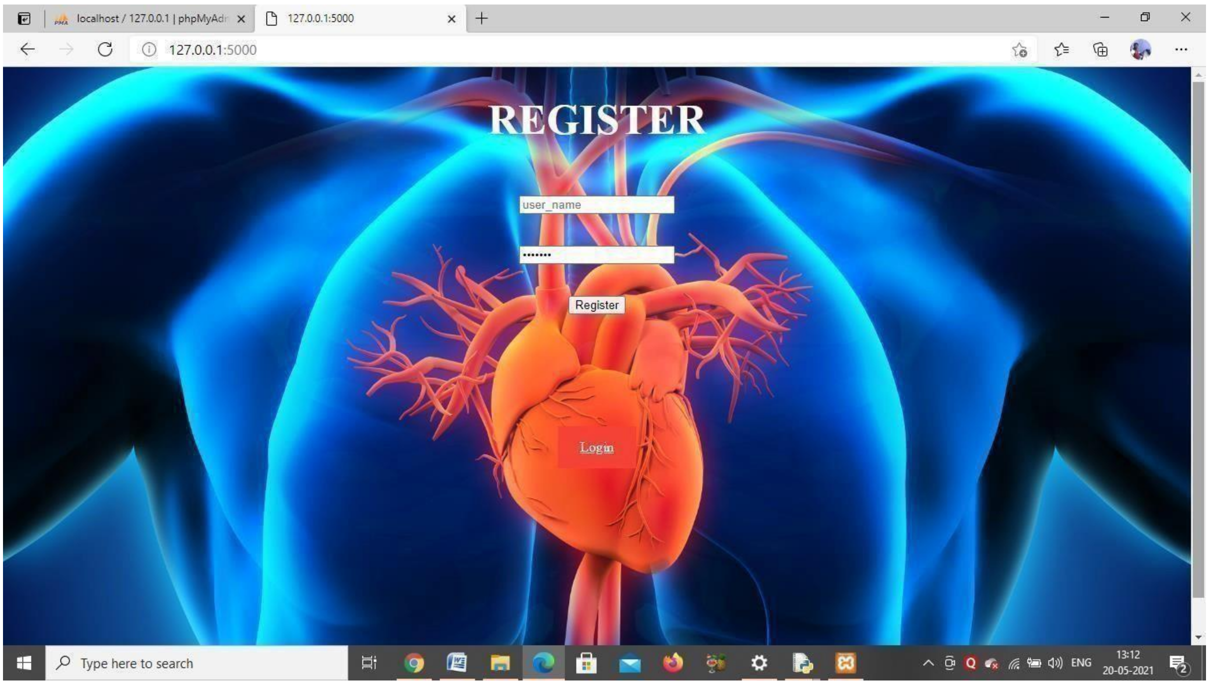
```
info = ["age","1: male, 0: female","chest pain type, 1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic","resting blood pressure"," serum cholestoral in mg/dl","fasting blood sugar > 120 mg/dl","resting electrocardiographic results (values 0,1,2)"," maximum heart rate achieved","exercise induced angina","oldpeak = ST depression induced by exercise relative to rest","the slope of the peak exercise ST segment","number of major vessels (0-3) colored by flourosopy","thal: 3 = normal; 6 = fixed defect; 7 = reversable defect"]

for i in range(len(info)):
    print(data.columns[i]+":\t\t\t"+info[i])

print(data.columns)
print(data.head())
print(data.describe())
print(data.corr())
data = data.drop_duplicates()
print( data.shape )
print(data.isnull().sum())
data=data.dropna()
print(data.isnull().sum())

data["target"].value_counts().plot(kind="bar", color=["salmon","lightblue"])
plt.xlabel("0 = No Disease, 1 = Disease")
plt.title("Heart Disease")
plt.show()

# Create a plot of crosstab
pd.crosstab(data.target, data.sex).plot(kind="bar",
                                   figsize=(10,6),
                                   color=["salmon","lightblue"])
plt.title("Heart Disease Frequency for Sex")
plt.xlabel("0 = No Disease, 1 = Disease")
plt.legend(["Female","Male"])
plt.show()

y = data.target
```

```
# Create separate object for input features
X = data.drop('target', axis=1)

# Split X and y into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y,  test_size=0.2, random_state=0)
# Print number of observations in X_train, X_test, y_train, and y_test
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)



model1= LogisticRegression()
model2=RandomForestClassifier(random_state=285) #285,1673
model3= KNeighborsClassifier(n_neighbors=9)
model4=DecisionTreeClassifier()
model5= GaussianNB()
model6=SVC(kernel='linear',C=10 ,gamma=0.0009)
model7=XGBClassifier()

model1.fit(X_train, y_train)
model2.fit(X_train, y_train)
model3.fit(X_train, y_train)
model4.fit(X_train, y_train)
model5.fit(X_train, y_train)
model6.fit(X_train, y_train)
model7.fit(X_train,  y_train)


## Predict Test set results
y_pred1 = model1.predict(X_test)
y_pred2 = model2.predict(X_test)
y_pred3 = model3.predict(X_test)
y_pred4 = model4.predict(X_test)
y_pred5 = model5.predict(X_test)
y_pred6 = model6.predict(X_test)
y_pred7 = model7.predict(X_test)
acc1 = accuracy_score(y_test, y_pred1) ## get the accuracy on testing data
```

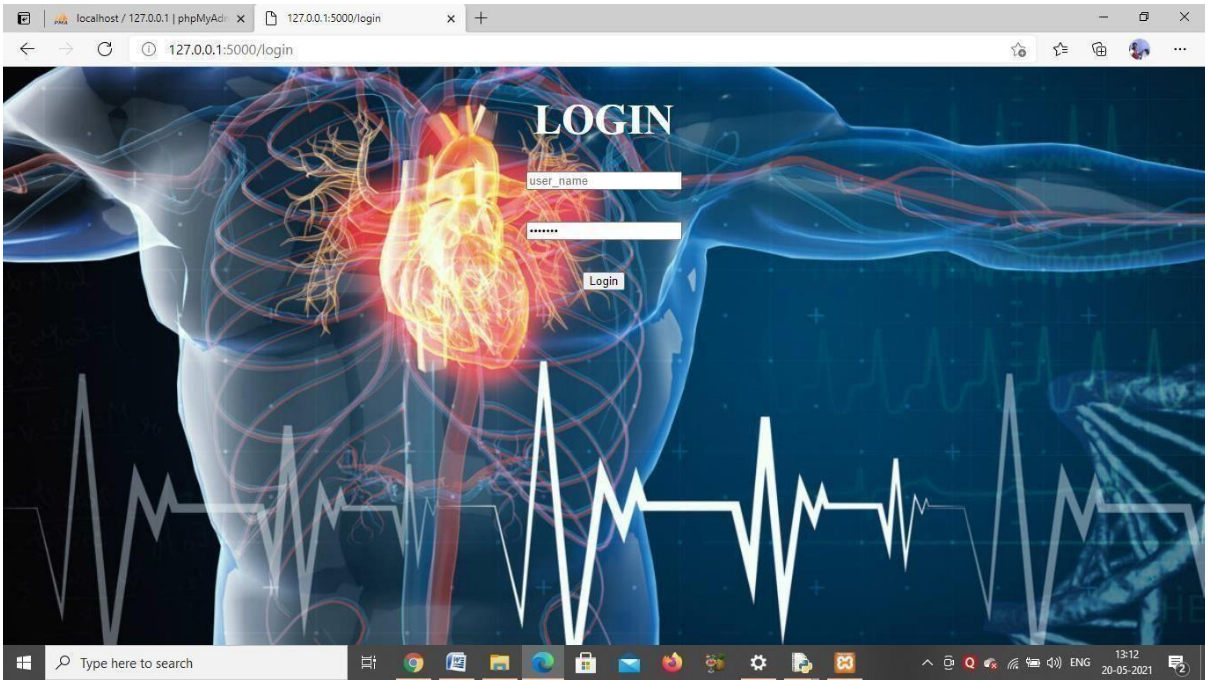
```
print("Accuracy of Logistic Regression is {:.2f}%".format(acc1*100))

acc2 = accuracy_score(y_test, y_pred2)  ## get the accuracy on testing data
print("Accuracy of RandomForestClassifier is {:.2f}%".format(acc2*100))

acc3 = accuracy_score(y_test, y_pred3)  ## get the accuracy on testing data
print("Accuracy of KNeighborsClassifier is {:.2f}%".format(acc3*100))

acc4 = accuracy_score(y_test, y_pred4)  ## get the accuracy on testing data
print("Accuracy of Decision Tree is {:.2f}%".format(acc4*100))

acc5 = accuracy_score(y_test, y_pred5) ## get the accuracy on testing data
print("Accuracy of GaussianNB is {:.2f}%".format(acc5*100))

acc6 = accuracy_score(y_test, y_pred6)  ## get the accuracy on testing data
print("Accuracy of SVC is {:.2f}%".format(acc6*100))

acc7 = accuracy_score(y_test, y_pred7)  ## get the accuracy on testing data
print("Accuracy of XGB Classifier is {:.2f}%".format(acc6*100)) #from sklearn.externals import
joblib
import joblib

# Save the model as a pickle in a file
joblib.dump(model2, 'heart_disease.pkl')

# Load the model from the file
final_model = joblib.load('heart_disease.pkl')

pred=final_model.predict(X_test)

acc = accuracy_score(y_test,pred)# get the accuracy on testing data
print("Final Model Accuracy is {:.2f}%".format(acc*100))
```

```
scores = [acc1,acc2,acc3,acc4,acc5,acc6,acc7]
algorithms = ["Logistic Regression","Random Forest","KNN","Decision Tree","Naive Bayes","SVC","XGB Classifier"]


sns.set(rc={'figure.figsize':(15,8)})
plt.xlabel("Algorithms")
plt.ylabel("Accuracy score")

sns.barplot(algorithms,scores)

plt.show()
```

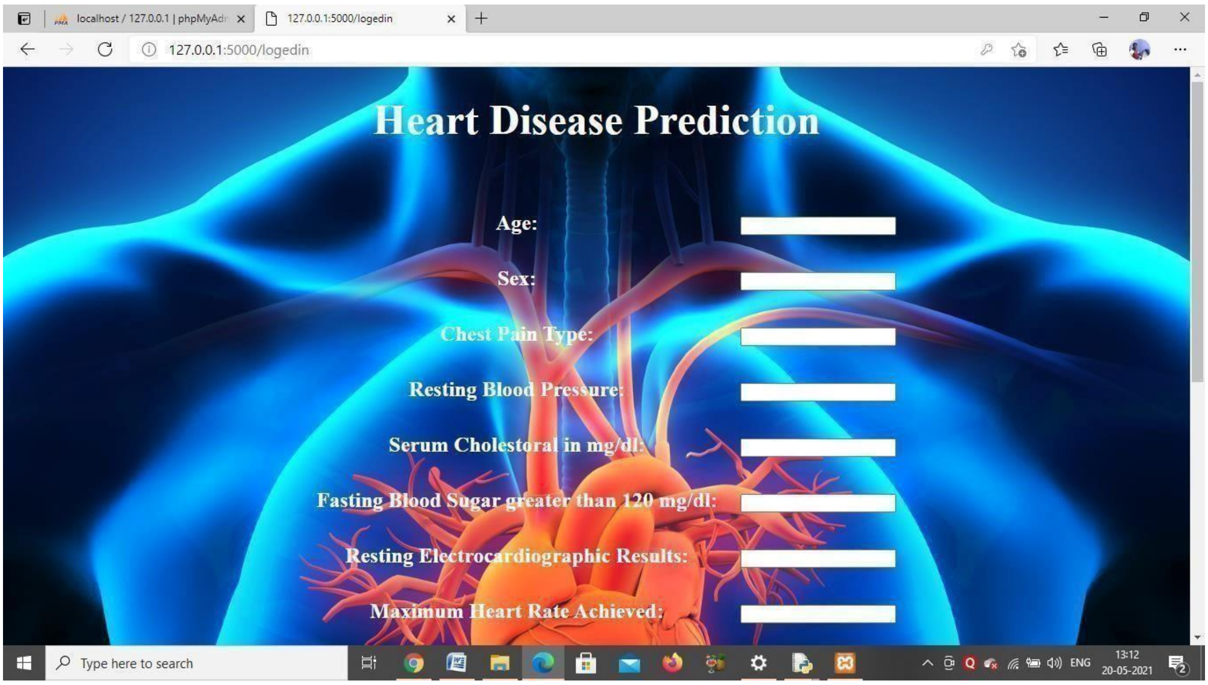
APPENDIX-B
SCREENSHOT



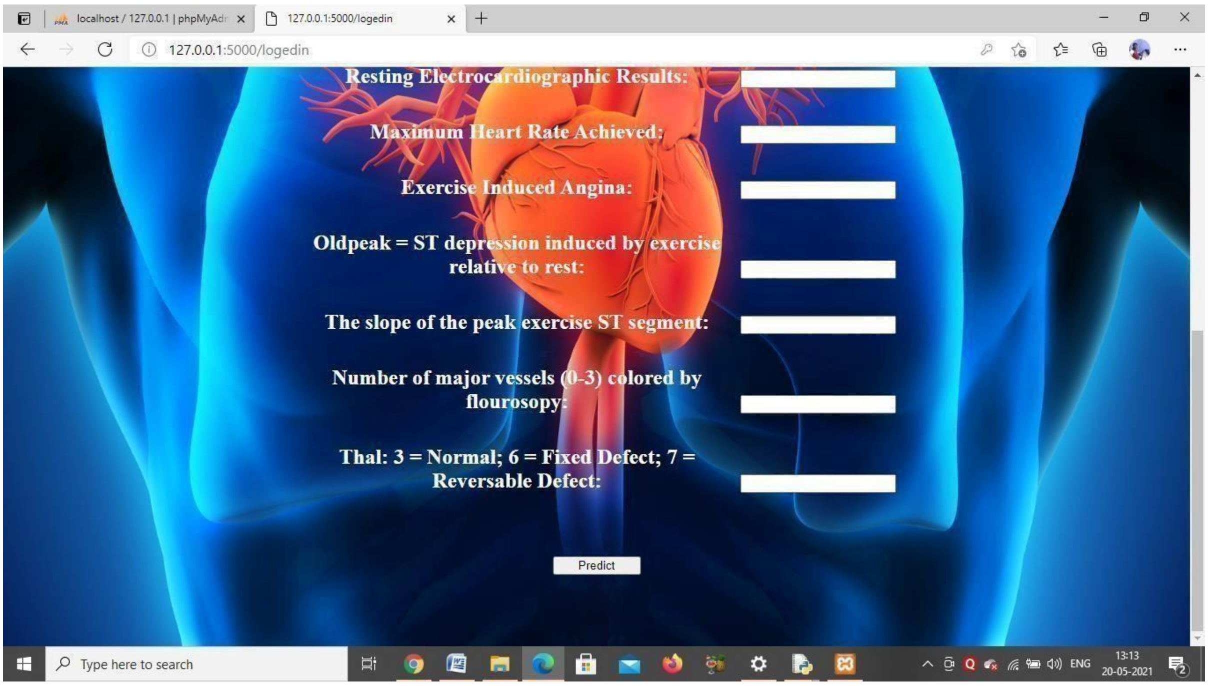
Register interface FIG. B1



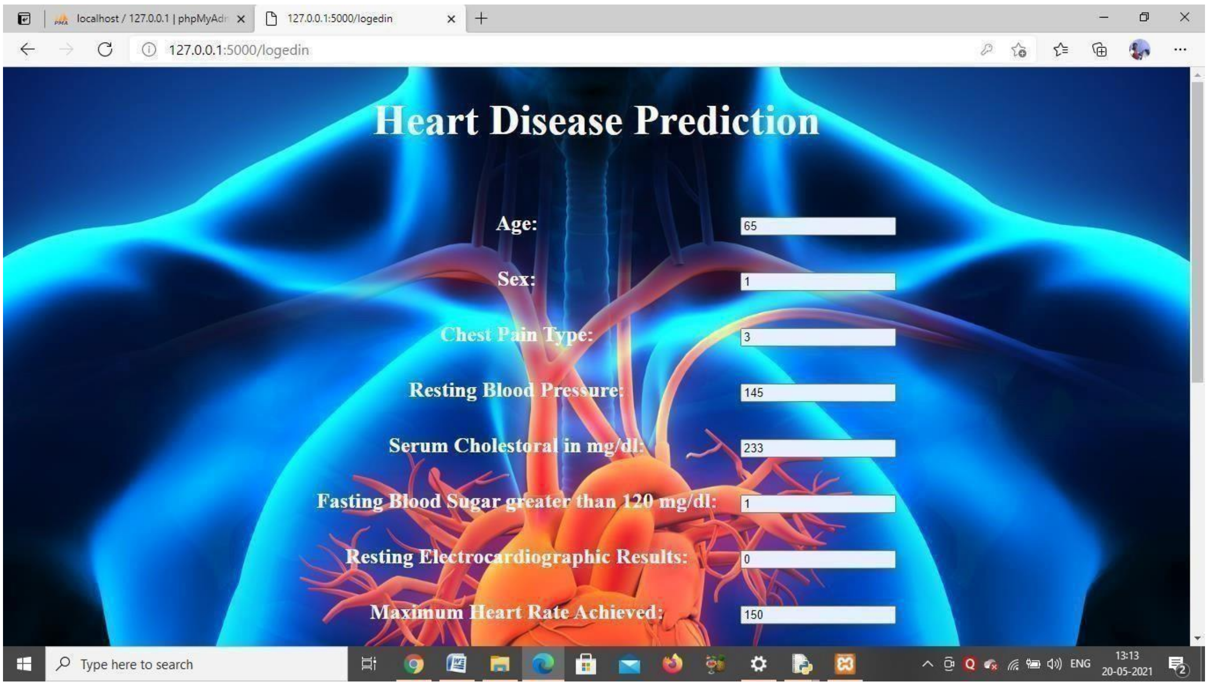
Login interface FIG. B2



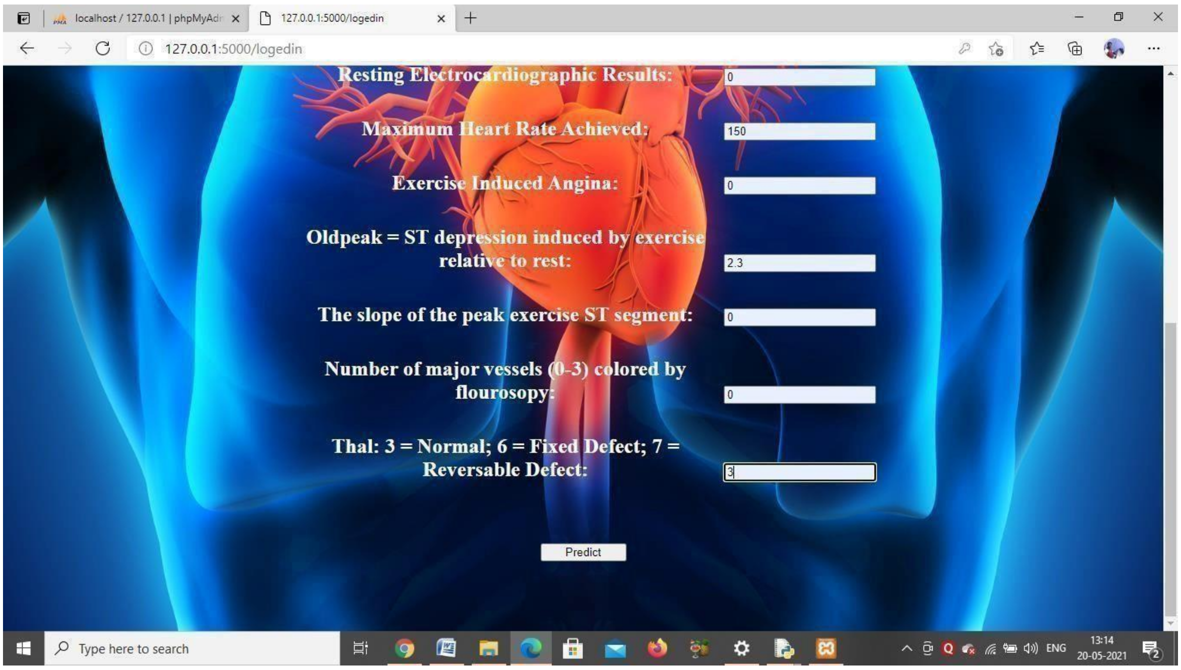
Prediction interface FIG. B3



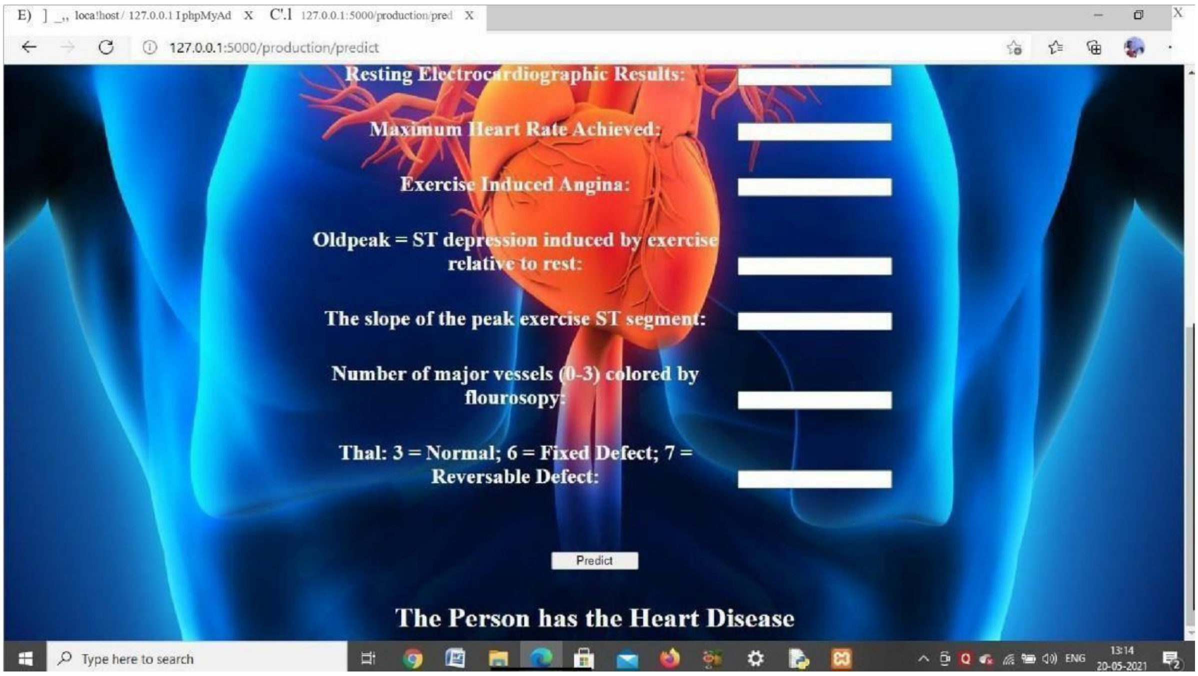
Prediction interface FIG. B4



FINALOUTPUT OFTHE PROJECT FIG. B5



FINALOUTPUT OFTHE PROJECT FIG. B6



FINAL OUTPUT OFTHE PROJECT FIG. B7

APPENDIX-C
ENCLOSURES

ORIGINALITY REPORT

16%	13%	9%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ijraset.com Internet Source	5%
2	Submitted to M S Ramaiah University of Applied Sciences Student Paper	2%
3	ijsrset.com Internet Source	1%
4	Submitted to University of Essex Student Paper	1%
5	123dok.net Internet Source	1%
6	Ritvik Gautam, MayankRaj, Abhishek Kumar, Kavita Saini. "Data Analysis on Covid-19", 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2021 Publication	1%
7	Submitted to Asia Pacific Instutute of Information Technology Student Paper	1%

8	www.dspace.dtu.ac.in:8080 Internet Source	1 %
9	www.ncbi.nlm.nih.gov Internet Source	1 %
10	www.ijrte.org Internet Source	1 %
11	www.researchsquare.com Internet Source	1 %
12	www.crn.com Internet Source	<1 %
13	Submitted to Kingston University Student Paper	<1 %
14	Submitted to University of Huddersfield Student Paper	<1 %
15	Monali Gulhane, T. Sajana. "A Machine Learning based Model for Disease Prediction", 2021 International Conference on Computing, Communication and Green Engineering (CCGE), 2021 Publication	<1 %
16	github.com Internet Source	<1 %
17	az659834.vo.msecnd.net Internet Source	<1 %

18	assets.researchsquare.com Internet Source	<1 %
19	Abderrahmane Ed-Daoudy, Khalil Maalmi. "Real-time machine learning for early detection of heart disease using big data approach", 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 2019 Publication	<1 %
20	ebin.pub Internet Source	<1 %
21	ijisrt.com Internet Source	<1 %
22	Sachin Kadian, Pratima Kumari, Shubhangi Shukla, Roger Narayan. "Recent advancements in machine learning enabled portable and wearable biosensors", Talanta Open, 2023 Publication	<1 %
23	"Artificial Intelligence Techniques for Advanced Computing Applications", Springer Science and Business Media LLC, 2021 Publication	<1 %
24	Santosh Kumar Satapathy, Harsh Makadiya, Mohammed Hamza Malik, Manav Dhami, Dhrupad Chauhan, Kunj Manavadariya. "Heart Disease: Automatic Prediction from	<1 %

the Numerical and Categorical Features by Machine Learning Methods", 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), 2023
Publication

25	Ubaida Fatima, Sadia Kiran, Muhammad Fouzan Akhter, Muhammad Kumail, Jaweria Sohail. "Unveiling the Optimal Approach for Credit Card Fraud Detection: A Thorough Analysis of Deep Learning and Machine Learning Methods", Research Square Platform LLC, 2024 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

International Journal of Research Publication and Reviews
(Open Access, Peer Reviewed, International Journal)
(A+ Grade, Impact Factor 5.536)
WWW.IJRPR.COM
ISSN 2582-7421
Sr. No: IJRPR 111729-1

Certificate of Acceptance & Publication

This certificate is awarded to "Lisha Shree N", and certifies the acceptance for publication of research paper entitled "Prediction of Cardiac Disease" in "International Journal of Research Publication and Reviews", Volume 5, Issue 1 .

IJRPR

Signed ____________________ *Date* 11-01-2024

Editor-in-Chief
International Journal of Research Publication and Reviews

International Journal of Research Publication and Reviews
(Open Access, Peer Reviewed, International Journal)
(A+ Grade, Impact Factor 5.536)
WWW.IJRPR.COM
ISSN 2582-7421
Sr. No: IJRPR 111729-2

Certificate of Acceptance & Publication

This certificate is awarded to "Himaja H B", and certifies the acceptance for publication of research paper entitled "Prediction of Cardiac Disease" in "International Journal of Research Publication and Reviews", Volume 5, Issue 1 .

IJRPR

Signed ____________________ *Date* 11-01-2024

Editor-in-Chief
International Journal of Research Publication and Reviews

International Journal of Research Publication and Reviews

(Open Access, Peer Reviewed, International Journal)

(A+ Grade, Impact Factor 5.536)

WWW.IJRPR.COM

ISSN 2582-7421

Sr. No: IJRPR 111729-3

Certificate of Acceptance & Publication

This certificate is awarded to "Pushpa Latha K", and certifies the acceptance for publication of research paper entitled "Prediction of Cardiac Disease" in "International Journal of Research Publication and Reviews", Volume 5, Issue 1 .

Signed ______________________ ***Date*** 11-01-2024

Editor-in-Chief
International Journal of Research Publication and Reviews

The project work carried out here is mapped to SDG-3 Good Health and Well Being.

The project work carried here contributes to the well-being of human society. This can be used for predicting cardiac disease so that medication can be started to avoid further consequences which might result in mortality.