An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases

Pulipati Madhurya[1], D Murali[2]
[1]*PG Student, CSE, Quba College of Engineering & Technology*
[2]*Associate professor, CSE, Quba College of Engineering & Technology*

Abstract*— Cardiovascular disorders are regarded as the most dangerous conditions, having the greatest fatality rate worldwide. They have become exceedingly common over time and are now overstretching national healthcare systems. High blood pressure, family history, stress, age, gender, cholesterol, BMI, and an unhealthy lifestyle are all key risk factors for cardiovascular disease. Researchers have proposed numerous ways for early diagnosis based on these criteria. However, due to the intrinsic criticality and life-threatening hazards of cardiovascular disorders, the accuracy of offered procedures and approaches need specific modifications. A Malc add framework is proposed in this study for the effective and precise prediction of cardiovascular disorders. The methodology, in particular, addresses missing values and data imbalances first.

Index Terms*— Cardiovascular disorders, Fatality rate, Risk factors, Early diagnosis, Healthcare systems, MaLCaDD framework, Prediction accuracy, Feature importance, Machine learning ensemble, Logistic Regression

I. INTRODUCTION

The current era's hectic pace leads to an unhealthy lifestyle that generates anxiety and despair to cope with these conditions, people tend to engage in excessive smoking, drinking, and drug use. All of them are the root causes of many severe diseases, such as cardiovascular disease and cancer According to the World Health Organization (WHO), cardiovascular diseases (CVDs) are the leading cause of death worldwide. CVDs account for over 31% of all fatalities worldwide Early detection of these disorders is critical so that precautionary actions can be implemented before something terrible occurs.

Cardiovascular Diseases (CVDs) are a group of conditions that affect the heart or blood vessels. Coronary Heart Disease, Stroke/Transient Ischemic Attack (TIA/ Mini Stroke), Peripheral Artery Disease, and Aortic Disease are the four major kinds of CVDs The actual origin of CVDs is still unknown; however, some risk factors for these diseases include high blood pressure, smoking, diabetes, body mass index (BMI),cholesterol, age, family history, and so on. These parameters vary from person to person. Age, gender, stress, and an unhealthy lifestyle are other key factors that contribute to CVDs The main challenge is to accurately predict these diseases in time so that mortality rates can be reduced through effective medication.

II .LITERATURE SURVEY

In literature a number of laptop mastering based totally analysis methods have been proposed through researchers to analysis HD. This lookup find out about current some current laptop getting to know based totally prognosis strategies in order to give an explanation for the vital of the proposed work. Detrano et al.developed HD classification gadget via the usage of laptop studying classification methods and the overall performance of the machine used to be 77% in phrases of accuracy. Cleveland dataset was once utilized with the technique of international evolutionary and with elements decision method.

In every other find out about Gudadhe et al. developed a prognosis machine the usage of multi-layer Perceptron and assist vector computer (SVM) algorithms for HD classification and finished accuracy 80.41%. Humar et al. designed HD classification machine with the aid of utilizing a neural community with the integration of Fuzzy logic. The classification device finished 87.4% accuracy. Result et al.

III. Project Description

3.1 About cardiovascular disease

A group of diseases that affect your heart and blood vessels is called cardiovascular disease. One or more parts of your heart and/or blood vessels can be affected by these diseases. A person may be asymptomatic (not feeling anything at all) or symptomatic (physically experiencing the disease). Cardiovascular disease includes problems with the heart or blood vessels, such as:

- A narrowing of the blood vessels throughout your body, including those in your heart and other organs.
- Issues with the heart and blood vessels are present at birth.
- Badly functioning heart valves
- Unsteady heartbeats

How common is cardiovascular disease?

The leading cause of death worldwide and in the United States is cardiovascular disease. Almost half of adults in the United States have some form of cardiovascular disease. People of all ages, genders, ethnicities, and socioeconomic classes are affected. One out of three ladies and individuals relegated female upon entering the world kicks the bucket from cardiovascular sickness.

3.3 Risk factors

There are many gamble factors for heart illnesses age, sex, tobacco use, actual latency, non-alcoholic greasy liver sickness, unnecessary liquor utilization, unfortunate eating routine, stoutness, hereditary inclination and family background of cardiovascular infection, raised circulatory strain (hypertension), raised glucose (diabetes mellitus), raised blood cholesterol (hyperlipidemia), undiscovered celiac sickness, psychosocial variables, neediness and low instructive status, air contamination, and unfortunate rest.

IV. SYSTEM ANALYSIS

4.1 EXISTING SYSYEM:

Based on risk factors, various methods for predicting CVDs have been proposed by researchers. Also, assortment of datasets have been involved by the scientists for the approval of their proposed approaches. Predictions were also made using k-means and clustering classifiers.

It can be used for a variety of healthcare and patient monitoring applications, including the prediction of heart disease.

4.1.1 DISADVANTAGES

- The accuracy is poor.
- Expensive time spent

4.2 PROPOSED SYSTEM

A MaLCaDD (Machine Learning based Cardiovascular Disease Diagnosis) framework is presented in this article. MaLCaDD intends to deal with missing values and data that isn't in balance in order to improve overall accuracy.

In this proposed system we are applying feature selection machine learning techniques they are KNN, SVM, Logistic Regression, MLP, Gaussian Naive Bayes, RNN-CNN, Random Forest Decision tree algorithms.

4.2.1 ADVANTAGES

- Precision is very high.
- Getting better results with fewer features is our primary objective.

V. SYSTEM DESIGN

A module diagram is a particular kind of diagram that shows the system's modules, components, and connections. Classes, interfaces, and other modules can all be included in modules, which are typically used to group functionality that is related. Module dependencies and code structure are frequently explained with the help of module diagrams. Developers can improve the code's overall structure by identifying potential issues and visualizing module dependencies. Additionally, member collaboration and communication can be enhanced by module diagrams.

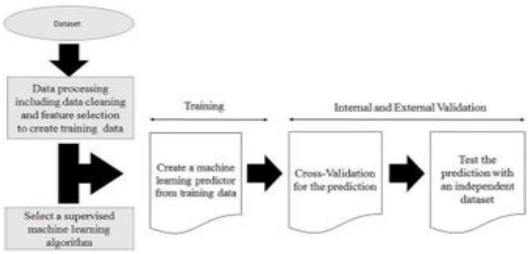


Fig. Modules

Data Pre-processing:

Pre-handling depicts the progressions made to our information before we feed it to the calculation. Data pre-processing is a method for transforming unclean data into clean data sets. All in all, information is constantly assembled from different sources in a crude express that blocks examination. A number of steps, including data cleaning, data transformation, and data reduction, are taken during data pre-processing to guarantee that the data are reliable, consistent, and suitable for analysis. It is a crucial step in the analysis of data because it increases the accuracy of the inferences made from it and improves the quality of the data.

Acquire the dataset:

Identify the data sources and gather the relevant data.

Ensure the data is in a format that can be processed (e.g., CSV, Excel, databases).Import necessary libraries:Import libraries such as pandas, NumPy, matplotlib, scikit-learn, etc., which are commonly used fordata pre-processing and analysis.

VI. PROJECT ANALYSIS

Analyzing various aspects of a project before, during, and after its completion is an essential step in project analysis. It distinguishes expected difficulties, dangers, and amazing open doors, permitting project directors and partners to pursue informed choices and make suitable moves.

Project analysis is important for the following key reasons:

Identifying project feasibility

Analyses help determine a project's feasibility by evaluating factors like market demand, budget constraints, technical requirements, and the availability of resources. It permits partners to arrive at informed conclusions about regardless of whether to continue with the undertaking.

VII. SYSTEM ARCHITECTURE

The conceptual model, also known as the system architecture, of a system determines its structure, behavior, and other characteristics. An architecture description is a formal description and representation of a system that is designed to make it easier to analyse its structures and behaviors.

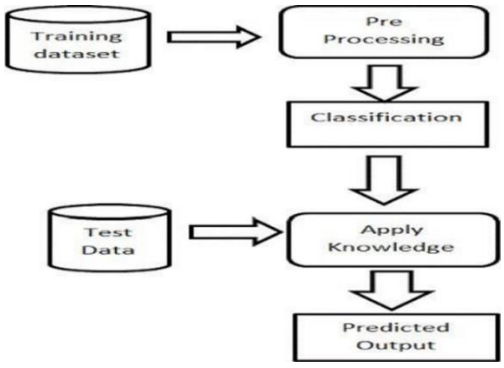


Fig: System Architecture

7.2 USE CASE DIAGRAM

A type of behavioral diagram known as a use case diagram is written in the Unified Modeling Language(UML) and is the result of a Use-case analysis. Its goal is to show a graphical overview of a system's functionality in terms of actors, their goals (shown as use cases), and any dependencies between those use cases. A use case diagram's primary goal is to indicate which actors use which system functions. Jobs of the entertainers in the framework can be portrayed.

Code

from flask import Flask, request, url_for, redirect, render templateimport joblib import pandas as pdimport sqlite3

app = Flask(name) model = joblib.

Load(open("model. Sav", "rb")) @app.route('/')

def hello world(): return render

template("home.html") @app.route('/logon')

def logon(): return

render_template('signup. html') @app.route('/login')

def login(): return

render_template('signin. html') @app.route("/signup")

def signup() username =

request.args.get('user',") name = request.args.get('name',"

) email = request.args.get('email',"

) number = request.args.get('mobile'

,")

password = request.args.get('password','

')con = sqlite3.connect('signup. dB')

cur = con.cursor()

cur. execute("insert into `info` (`user`,`email`, `password`,`mobile`,`name`)

S (?, ?, ?, ?, ?)",(username,email,password,number,name))con.commit() con.close()

return render_template("signin.html")

VIII SYSTEM TESTING

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

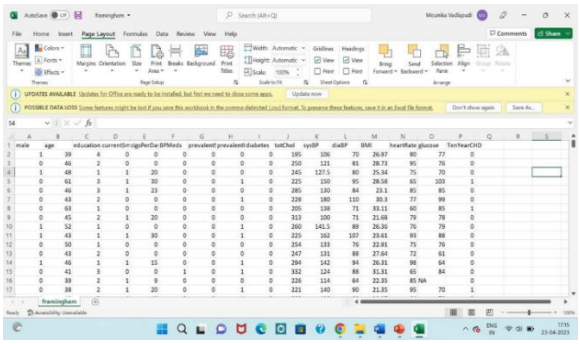
Test cases:

S.NO	Test case Description	Input given	Expected Output	Actual Output	Result
1	Upload dataset	Upload dataset	Dataset imported	imported	pass
2	Run algorithm	Click on run algorithm	Algorithm Executed and got accuracy	Got accuracy	pass
3	Click on graph	Click on graph to View accuracy among algorithms	Graph generated	Graph Generated	pass
4	Predict Result	In this we predict output results based on test data	Predict result	Predicted	pass

IX. IMPLEMENTATION

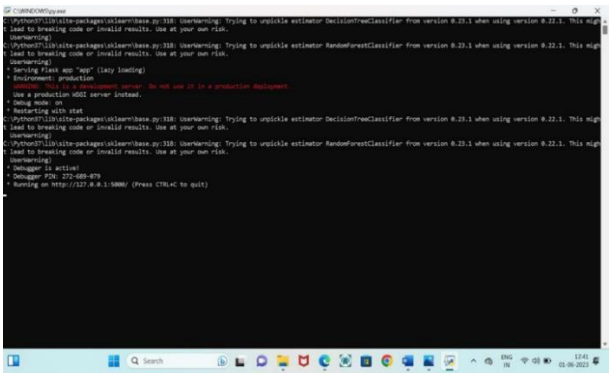
9.1 Dataset:

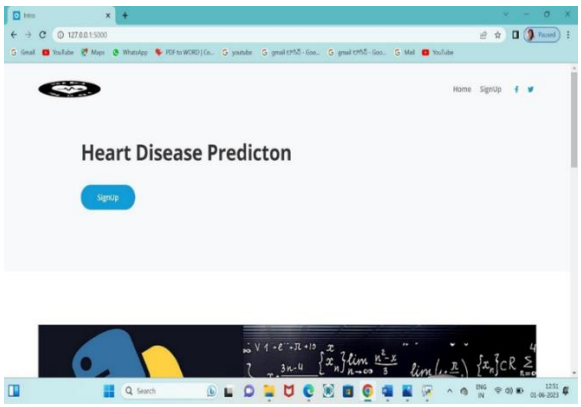
Below screen shots showing dataset columns and its values



In above dataset first line contains section names and different lines contains values as 0 or 1 and in the event that patient is under Cardiovascular illness, its segment worth will be 1 else 0 and in last section contains class mark as 0 or 1 where 0 methods patient record is typical and 1 method patient record contains Cardiovascular disease.

X SCREENSHOTS





X.CONCLUSION

The MaLCaDD platform for the early prediction and diagnosis of cardiovascular diseases is discussed in this project. The framework is broken up into four main phases, the first of which uses the mean replacement method to manage missing values. The Synthetic Minority Oversampling Technique (SMOTE) is utilized in the second phase to address the imbalance in the data. In the third phase, features are selected using the feature significance technique.

KNN Classifier, Logistic Regression, Decision Tree Classifier, Ensemble of LR and Decision Tree, Random Forest, and SVM are all suggested for improved prediction. The Python-based MaLCaDD implementation can be downloaded from the GitHub repository.

REFERENCES

[1] Organization Mondale de la santo. 2017). Diseases of the heart (CVDs) Online]. Available: https://www.who.int/wellbeing points/cardiovascular diseases

[2] E. J. Benjamin et al., " Statistics on heart disease and stroke—update for 2019: The American Heart Association's report, Circulation, vol. 139, no. 10, pp. Mar. 2019, e56–e528, Doi: 10.1161/CIR.0000000000000659.

[3] Korea Statistics 2018). Statistics on Causes of Death in 2018 [Online]. Available: http://kostat.go.kr/portal/eng/pressReleases/8/index.board?bmode=read&b Seq=&aSeq=378787 2017).Diseases of the heart (CVDs) Online]. Available:

[4] P. Greenland, J. S. Alpert, G. A. Beller, E. J. Benjamin, M. J. Bud off, Z. A. Fayad, E. Foster, M. A. Hlatky, J. M. Hodgson, F. G. Kushner, M. S. Lauer, L. J. Shaw, S.

[5] C. Smith, A. J. Taylor, W. S. Weintraub,and N. A report on practice guidelines from the American Heart Association/American College ofCardiology Foundation, Circulation, vol. 122, no. 25, pp. Dec. 2010, e584–e636, doi:10.1161/CIR.0b013e3182051b4c.