Project Report

on

"Explainable AI in Cardiovascular Diseases"

Submitted by

Project Members

Sakshi Kothari	PRN Number: 1032200701
Ritu Karnawat	PRN Number: 1032192159
Veda Dhapke	PRN Number: 1032191389
Utkarsh Mukkawar	PRN Number: 1032192230

Under the Internal Guidance of

Prof. Nitin Pise

School of Computer Engineering and Technology
MIT World Peace University, Kothrud,
Pune 411 038, Maharashtra - India

2022-2023

SCHOOL OF COMPUTER ENGINEERING AND TECHNOLOGY

C E R T I F I C A T E

This is to certify that,

Sakshi Kothari,

Ritu Karnawat,

Veda Dhapke

UtkarshMukkawar

of BTech.(Computer Science & Engineering) have completed their project titled "*Explainable AI in Cardiovascular Diseases*" and have submitted this Capstone Project Report towards fulfillment of the requirement for the Degree-Bachelor of Computer Science & Engineering (BTech-CSE) for the academic year 2022-2023.

[Prof. Nitin Pise]
Project Guide
School of CET
MIT World Peace University, Pune

[Dr. Vrushali Kulkarni]
Program Head
School of CET
MIT World Peace University, Pune

Internal Examiner: ______________________

External Examiner: ______________________

Date:

Acknowledgement

We are highly indebted and would like to acknowledge and give our warmest thanks to our guide Prof.Nitin Pise who made the accomplishment of this project possible. His valuable guidance , support and advice carried us through all the steps of the project.

We would also like to thank the School of Computer Engineering and Technology(SCET) and the head of school Dr Vrushali Kulkarni for providing this valuable opportunity.

We also acknowledge the reviewers' anonymity and appreciate their attentive reading of our article as well as their numerous useful remarks and recommendations.

Name of the Students:

Sakshi Kothari (1032200701)

Ritu Karnawat (1032192159)

Veda Dhapke(1032191389)

Utkarsh Mukkawar(1032192230)

Abstract

Due to the availability of enormous data files and increased computing power in the past decade, machine learning algorithms have achieved impressive performance in a wide range of jobs.Applications like image identification, predictive analysis, medical diagnosis, extraction, speech recognition, and many more are examples of how machine learning has advanced.Even though machine learning has advanced significantly, the problem is the lack of transparency and explicability. In many implementations, such as those in healthcare and finance, where the approach's explainability is a matter of trust, the absence of the aforementioned criteria is a significant problem.

Acknowledging these issues, Explainable artificial intelligence(XAI) has turned into a domain of interest.By doing this, we may further dissect these technologies to determine whether there is a method to get the same predictions faster or with better algorithms.

In this study, we use a variety of interpretability methodologies to evaluate these technologies on a bigger dataset and compare several machine learning algorithms with different levels of accuracy.Our research illustrates why explainability tactics should be adopted when deploying AI systems in healthcare to establish credibility by using instances taken from the heart disorders dataset.

Keywords: Explainable AI, Healthcare, Heart disease,, LIME, SHAP, Machine Learning Techniques.

List of Figures

List of Tables

Contents

Chapter 1
Introduction

1.1 Project Statement

Explainable AI in Cardiovascular Disease.
Through this study, we focus upon how several ML techniques such as Random Forest, Decision Tree and Gradient boosted tree predicts the end result of CVD dataset through various XAI techniques

1.2 Project Domain

Artificial Intelligence, Machine Learning ,Explainable Artificial Intelligence

1.3 Motivation

According to estimates, 17.9 million people worldwide die from cardiovascular disorders each year (Reference: WHO statistics). Heart attacks and strokes are to blame for four out of every five CVD fatalities. Our project's primary motivation is to lessen the mortality rate from cardiovascular diseases by anticipating them well in advance so that individuals can take preventative measures.

1.4 Project Introduction and aim

Nowadays, age is not a factor for the increase in death rate due to various diseases. Lifestyle changes, pollution and certain other factors are responsible for this increase.Particularly speaking about cardiovascular diseases and stroke,bad food, lack of exercise, smoking, alcohol consumption, and obesity, are among the factors responsible for CVD's.

For people and medical professionals to make decisions and be sure that the results are reliable, explanation of the fundamental reasoning is vital for CVDs and other healthcare applications. Lack of explainability in machine learning limits the expansive application of AI. If AI is unable to explain how it predicts outcomes in the healthcare industry, the danger of making a bad choice may outweigh its benefits of precision and speed, which would restrict its use. Standard tools must be created in order to tackle these problems and use AI to predict CVDs in advance and take preventative measures. Explainable AI is one such tool (XAI). It is possible to use XAI to explain how and why a person is predicted to develop cardiac disease in the near future. As a result, medical professionals would have more confidence in AI, which would allow them to treat patients earlier and lower the death rates from CVDs.

Health information collected at a user level can also be shared with clinicians for further diagnosis and together with AI can be used in health screening, early diagnosis of diseases, and treatment plan selection. In the healthcare domain, the ethical issue of transparency associated with AI and the lack of trust in the black-box operation of AI systems creates the need for AI models that can be explained . Further XAI integration into intelligent healthcare systems is possible for more disorders including cancer, neurology, etc. These technologies can be used to diagnose serious illnesses and choose the best course of therapy. Our research aims to describe the many XAI methods used on the Kaggle dataset for cardiovascular disorders, including SHapley and LIME.

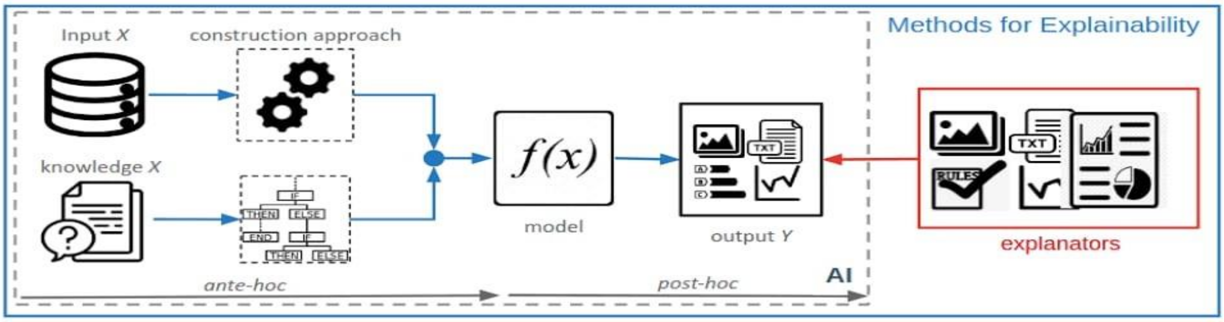


Fig.1.4.1: Brief Diagram of Explainable AI

Chapter 2
Literature Survey

The themes of traditional risk prediction models, artificial intelligence (AI) used to recognise CVD in various types of imaging data, or AI for the purposes of drug discovery and electrocardiogram-based prediction have all received a lot of attention recently. Reviews that take into account the molecular and genetic elements of CVDs have also been published recently, with the majority of them concentrating on detailing the intricate processes that underlie disease. These reviews take incident CVD into account as opposed to recurring occurrences. Here, however, we concentrate on recent developments in risk prediction of recurrent cardiovascular events based on clinical and molecular data and highlight the application of AI as a tool for forecasting risk and developing individualized treatment plans. Because performance measures are frequently used to objectively evaluate prediction models,Finally, we discuss how explainable AI (XAI) may be a key component of future transparent and dependable clinical decision systems that are based on both clinical and molecular data. These issues in risk prediction can now be handled with AI.

One of the most common model-agnostic approaches is called Local Interpretable Model-Agnostic Explanation (LIME), which is a framework for explaining predictions by quantifying the contributions of all the factors involved in calculating predictions. Researchers used LIME to describe how Recurrent Neural Networks (RNNs) anticipate heart failure, and their explanations helped pinpoint the most common illnesses that raise a person's risk of developing heart failure, such as diabetes, anemia, and renal failure. Anchors and Shapley values are two further model-independent XAI techniques that have been created and are used in the healthcare sector.

[6]We have looked at some of the main arguments put forth in the literature against the pervasive application of XAI methods in medical settings. Some of the concerns and arguments put up in response to them can actually be reframed to apply to applications outside of the healthcare industry. However, we should be aware of how the discussion is affected by the increased stakes involved in safety-critical AI applications in the healthcare industry. It is important to note that downgrading explainable medical AI's usefulness would mean suppressing efforts to obtain human oversight over a cutting-edge technology that is still in its

infancy but has the potential to be quickly and widely implemented, potentially affecting a large portion of the population.XAI is not a panacea for all of AI's issues, and it is especially not a replacement for meticulous model performance analysis. However, there are a number of situations where XAI technologies can be helpful. More black boxes won't stop the loss of human control in a rapidly digitizing society since human monitoring of machines is still a crucial principle. Enhancing human capabilities has been claimed to be better to automation since the latter causes deskilling, detachment, and increases in unemployment.The prospect of a human-machine synergy, or a means for people to interact with AI, is assumed by augmentation. The European Union and other public organizations are debating rules for AI goods. Although the XAI field may not yet be mature enough for XAI techniques to be included into rules as hard requirements, this does not mean that XAI approaches should be ignored or put on the back burner. In order to exert an additional layer of control over AI products, the aforementioned strategies can be used; however, as with any technology, users should be aware of any potential failure modes. In terms of rules,Allowing the usage of XAI techniques while requesting an explanation as to why a particular technique was selected could be useful.

Sr. No.	Author's Name	Paper Name	Year	Technique Used	Research Gap
1	Urja Pawar,Donna O'Shea,Susan Rea,Ruairi O'Reilly	Explainable AI in Healthcare	June 2020	Local Interpretable Model-Agnostic Explanation (LIME)	
2	Giovanni Cina, Tabea Rober, Rob Goedhar, lker Birbil	Why we do need Explainable AI for Healthcare	30 Jun 2022	Adversarial attacks on SHAP (one of the techniques offering individual-level explanations) suggest that out-of-distribution samples should not be employed in approximate SHAP calculations.	XAI techniques will almost never offer a one-size-fits-all solution.

3	Guang Yang ,Qinghao Ye ,Jun Xia	Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion:	2022	Model-specific local XAI, Model-agnostic global XAI	The design of a computing system for the processing, storage and exchange of EHRs and other critical health data remains a problem
4	*Vishal Sharma, Piyush, Samarth Chhatwal, Bipin Singh*	An Explainable Artificial Intelligence based Prospective Framework for COVID-19 Risk Prediction	5 March 2021	Grid search cross validation(for searching and selecting hyperparameters), LIME explanation for the prediction made by the Custom 3-layer CNN model, XGBoost model.	Since this is a comparative study, every provided technique lacks something which is an advantageous attribute of the other methods.
5	Salih Sarp, Murat Kuzlu, Emmanuel Wilson, Umit Cali, Ozgur Guler	The Enlightening Role of Explainable Artificial Intelligence in Chronic Wound Classification	11 June 2021	Local Interpretable Model-Agnostic Explanations (LIME) to find the statistical connection between input and model prediction, XAI methods	Performance metric evaluation of the model on diabetic wounds indicate that the model has limitations with feature identification for this wound type.
6	Thomas Plouga, Søren Holm	The four dimensions of contestable AI diagnostics - A patient-centric approach to explainable AI	2020	Contestable AI in AI Diagonistic	Contestability is domain-specific, Contestability is a weaker requirement than simulatability
7	Annie M. Westerlund, Johann S. Hawe, Matthias Heinig, Heribert Schunkert	Risk Prediction of Cardiovascular Events by Exploration of Molecular Data with Explainable Artificial Intelligence	24 September 2021	Cox PH(Proportional Hazards), Principal Component Analysis(PCA), Logistic Regression	Polygenic risk scores(PRS) for CAD doesn't enhance recurrent even predicition.
8	Devam Dave , Het Naik , Smiti Singhal, and Pankesh Patel	Explainable AI meets Healthcare: A Study on Heart Disease Dataset	6 Nov 2020	Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), Anchors, Counterfactuals ,Contrastive Explanation Methods, Kernel Shapley	As it is a comparative study, all these techniques lack something which is the major attribute of another one of these techniques

9	Marzyeh Ghassemi, Luke Oakden-Rayner, Andrew L Beam	The false hope of current approaches to explainable artificial intelligence in health care	November 2021	SHAP, LIME	Explainability methods cannot yet provide reassurance that an individual decision is correct, increase trust among users, nor justify the acceptance of AI recommendations in clinical practice.
10	Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter and Lalana Kagal	Explaining Explanations: An Overview of Interpretability of Machine Learning	2018	Scripted conversations Attention-based Disentangled rep. Human evaluation	the various approaches taken to address different facets of explainability are siloed. Work in the explainability space tends to advance a particular category of technique, with comparatively little attention given to approaches that merge different categories of techniques to achieve more effective explanation.
11	Weina Jin , Xiaoxiao Li, Mostafa Fatehi, Ghassan Hamarneh	Guidelines and evaluation for clinical explainable AI on medical image analysis	16 Feb 2022	Clinical Explainable AI ,Multi-modal medical imaging	Evaluated heatmap methods failed to meet the Clinical XAI Guidelines
12	Md. Sarwar Kamal , Nilanjan Dey, Syed Irtija Hasan , and KC Santosh	Explainable AI for Glaucoma Prediction Analysis to Understand Risk Factors in Treatment Planning	2022	SNN, ANFIS,	ANFIS with SNN provides better performance than CNN all structures because ANFIS & SNN process both medical numerical data and glaucoma images. CNN, on the other hand, only processes image data

Table 2.1. Literature Survey of relevant Research Papers

Chapter 3
Problem Statement

3.1 Project Scope

The scope of the project is to consider certain factors like glucose, cholesterol, alcohol consumption, smoking , blood pressure etc and then predict if the person will have the cardiovascular diseases in the future or not based on machine learning algorithms and then explain the predictions using Explainable AI algorithms like LIME and SHapley.

3.2 Project Limitations

As you can see ahead , it is difficult to characterize or analyze results we received from SHAP as the whole basis of SHAP is feature extraction , which for explaining the prediction considers only two values, those are Cholesterol and High Blood pressure.

3.3 Project Objectives

The objectives of the project are to elucidate the predictions stated by the machine learning techniques. The explainability given by both explainable techniques(LIME and SHapley) are then analyzed and different testing and training percentages are applied to get the best accuracy.

Chapter 4
Project Requirements

4.1 Resources

In this project we will be using Google Colab and Jupyter notebook where we would work with various python libraries and machine learning methods and XAI techniques.

4.2 Reusable Software Components

We have implemented data preprocessing like checking null values, if any then replacing them with mean /median values. In our CVD dataset, the age factor was given in the number of days , we then converted it into years and this preprocessed data is further used ahead for predictions and explainability.

4.3 Software and Hardware Requirements

- Hardware Requirements:
    - Intel i3 8th generation and above
    - 4GB Ram
    - 500 GB SSD/ HDD

- Software Requirements:
    - 64 bit OS
    - Windows 7 and above
    - Python 3+
    - Google colab
    - Jupyter Notebook
    - Python Libraries

4.4 Requirements Rationale

Sr.no	Requirement	Rationale
1	All the algorithms must provide a graphical output	We imported library like sns
2	In-depth Explanation	We also checked and compared single data entries.

Table 4.4.1. Requirements rationale

4.5 Project Risk factors in Table format

Factor	Cause	Impact	Chances
Frequent software crashes	Data size (i.e. 70,000)	Might affect the speed we are doing the project with	Low
Shapley algorithm	Only takes important characteristics	we can't get overall reason for the prediction	Medium
Types of results	Both of the algorithms have different kinds of results	As a comparison study these factors can slow us down	High

Table 4.5.2. Risk factors

4.6 Functional Specifications:

4.6.1 Interfaces

- **External interfaces required(functions)**

    1. Random Forest:

```
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.8, random_state=42)


rf = RandomForestRegressor(n_estimators = 200, random_state = 42)
rf
```

2. DecisionTree:

```
from sklearn import tree

X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.8, random_state=42)
clf = tree.DecisionTreeClassifier(max_depth=3, criterion='entropy')
clf
```

3. Gradient Boosted:

```
from sklearn.ensemble import GradientBoostingClassifier

X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.8, random_state=42)
gb_clf2 = GradientBoostingClassifier(n_estimators=500, learning_rate=0.7, max_features=4,
max_depth=4, random_state=0)
gb_clf2
```

- **Internal interfaces required**

1. Lime Explainer :

```
predict_fn_rf = lambda x:"model".predict_proba(x).astype(float)
X = X_train.values
explainer = lime.lime_tabular.LimeTabularExplainer(X,feature_names =
X_train.columns,class_names=['Has CVD','Doesnt have CVD'],kernel_width=5)
```

2. Shap Explainer :

```
explainer = shap.TreeExplainer(model)
choosen_instance = X_test.iloc[[669]] 11/15/22, 12:35 PM Shap_randomforest
shap_values = explainer.shap_values(choosen_instance)
```

- **Graphical User Interfaces**
  1. GUI For Lime:

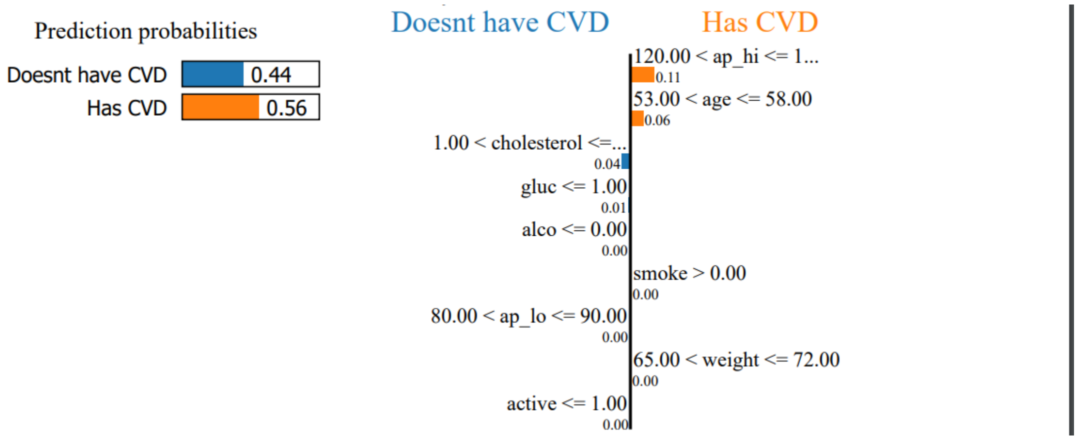


Fig.4.6.1.1. Graphical User Interface for LIME

  2. GUI For Shap:

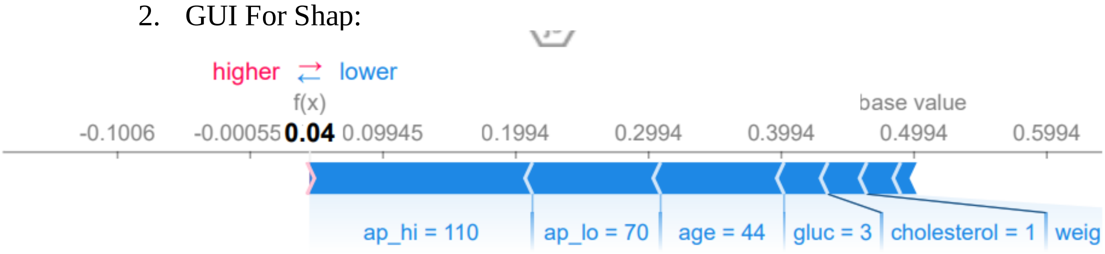


Fig. 4.6.1.2. Graphical User Interface for SHAP

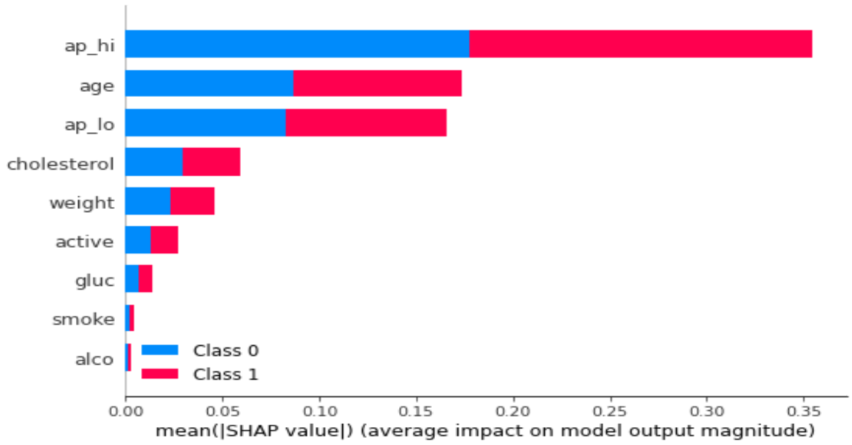


Fig. 4.6.1.3. Feature impact

Chapter 5
System Analysis Proposed Architecture

5.1. Design Consideration

- Maintainability: As the application has a huge database and the maintenance of the database will be difficult as the application will split the database on regular intervals and hence maintainability is one of the most important design considerations of this application.
- Reusability: As the code can be used for different models/explainers and the data of faculties and research papers can be used for different erp models, the application should be reusable.

5.2. Assumptions and Dependencies

- The main assumption about this project is that it will explain the values by which a specific machine learning model created its prediction, in addition to highlighting the significant features that have an impact.
- Also the other assumption is that the explainers provide a range of graphs , heat maps ,etc for better understanding.

5.3. General Constraints

- An explanation needs two sorts of inputs in order to function: the trained model and data entered for the test side of the data.
- However, SHAP only uses some of the characteristics from these inputs, not all of them, making it challenging to examine through.

5.4. Block Diagram

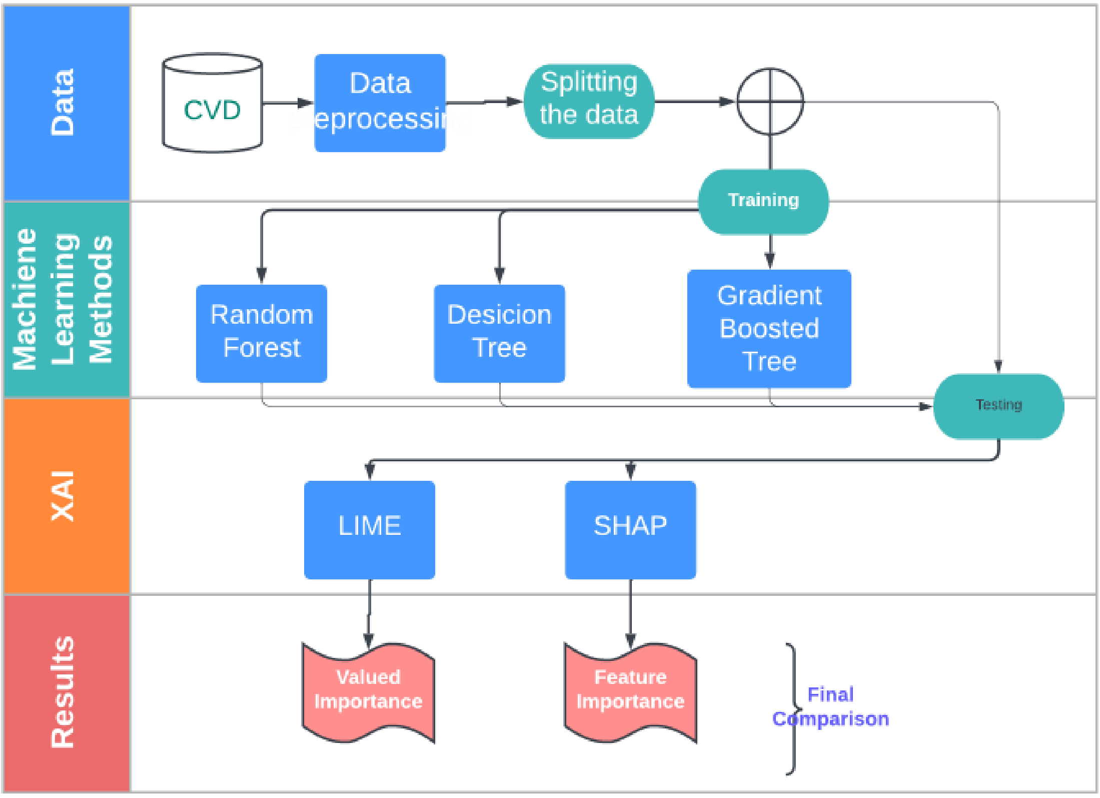


Fig. 5.4.1. Block Diagram for the system

5.5. System Architecture

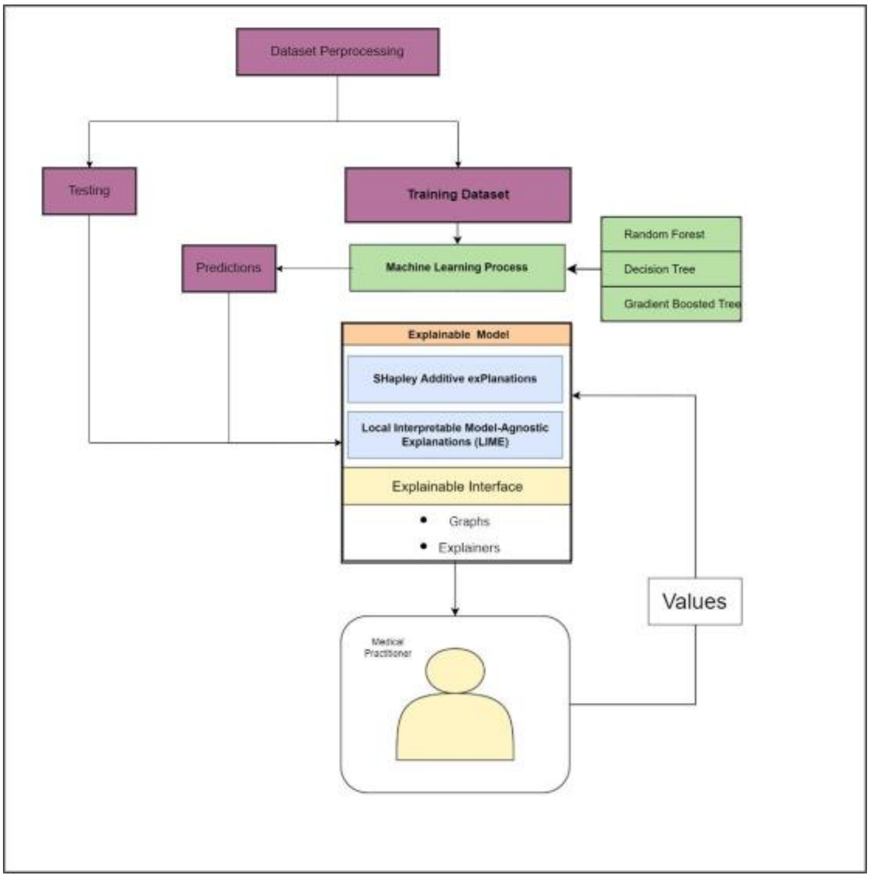


Fig. 5.5.1. System Architecture Diagram

5.6. Modules of the Project

Number of Modules :

- Data pre-processing
- Machine Learning Process
- Explainable AI

1) Data Preprocessing

Here we apply various preprocessing ways to prune the dataset so that it can be more optimal for usage.

2) Machine Learning Process

In order to examine the database and select the predictions with the highest accuracy, we used a variety of techniques, including Bayesian, KNN, Decision tree, Random forest, Gradient Boosted tree, etc. While doing so, we discovered that the Gradient boosted tree

(0.7902142857142858) provided the highest level of accuracy, followed by Decisions tree, Random Forest, KNN, Bayesian, etc. So, we decided to go with the top three.

3) Explainability of AI

For this project we are using methods like SHap (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) on our dataset so we can understand the way a machine predicts data. These predictions would be analyzed through various factors such as trees, Graphs , Heatmaps, etc.

5.7. Low Level Design

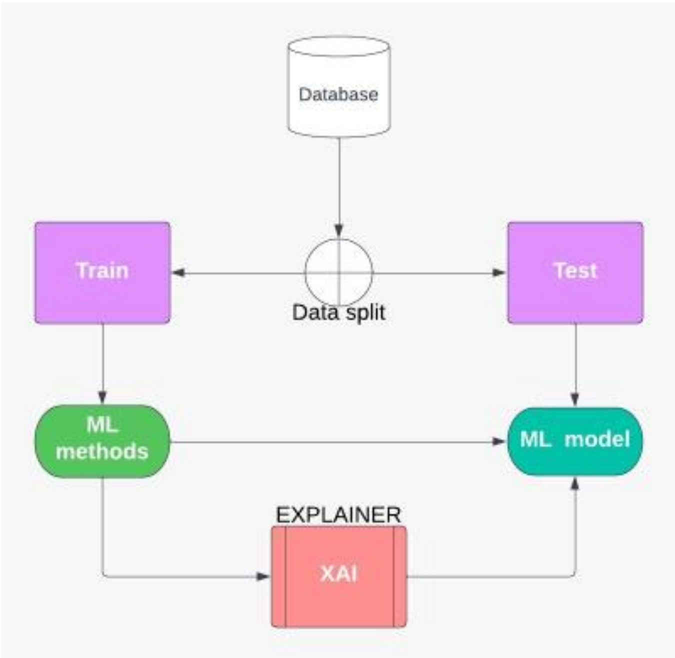


Fig. 5.7.1. Low level diagram

5.8. UML Diagrams/Agile Framework

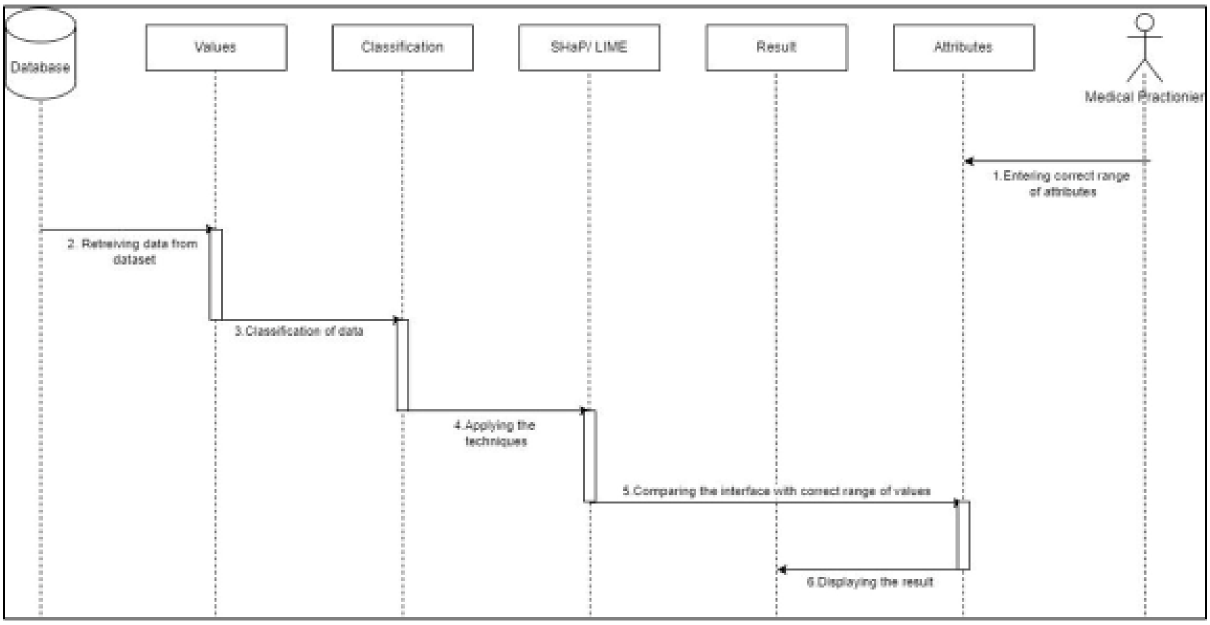
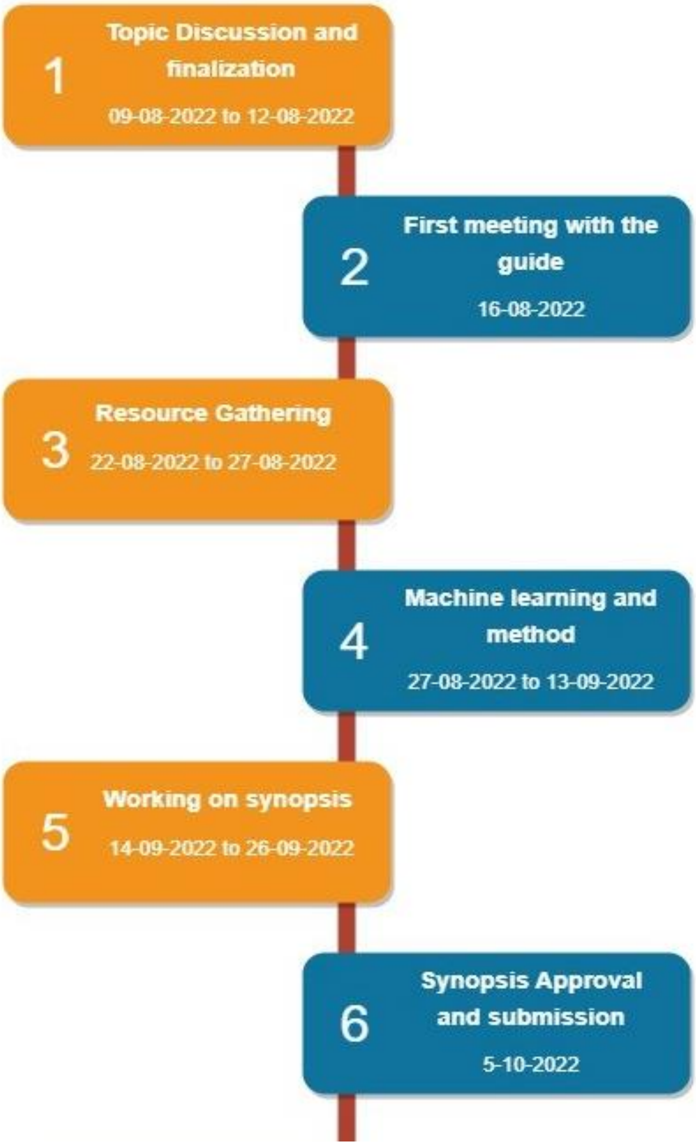


Fig. 5.8.1. Sequence Diagram

Chapter 6
Project Plan

Timeline chart for the entire SDLC of our project:



6.1. Timeline Diagram of the project

Chapter 7
Implementation

7.1. Methodology

In this study, we classified our dataset using three machine learning techniques, i.e., Random Forest, Gradient Boosted Tree, and Decision Tree. Moving forward, we elaborated the given methods using Explainable AI technique- LIME, which builds a local interpretable model with the intent of calculating feature values.

Provided below is the LIME explainer code we have implemented in our analysis:

predict_fn_rf = lambda x: random_forest.predict_proba(x).astype(float)

X = X_train.values

explainer=lime.lime_tabular.LimeTabularExplainer(X,feature_names=X_train.columns,class_names=['Doesn't have CVD','Has CVD'],kernel_width=5)

Herein, we have initially calculated the prediction probability of our data using the Random Forest predict_proba method. This method estimates class probabilities for the input x. These probabilities are calculated as average predicted class probabilities of the trees in the forest. We have stored the trained values in a variable and then written the explained code which uses the lime_tabular.LimeTabularExplainer method which takes the parameters as x_train.columns and gives class names as "Doesn't have CVD" and "Has CVD".

7.2. Algorithm

- **Local Interpretable Model-agnostic Explanations (LIME)**

LIME which is Local Interpretable Model-agnostic Explanations is an explainable AI technique which helps in throwing light on a machine learning model and making every prediction's specific inference understandable.Due to the fact that it describes the classifier for a specific single instance, this technique is appropriate for local explanations.

LIME modifies the input data to produce or generate a sequence of fictional data with just a minimal amount of the original properties retained. As a result, multiple replicas of original text are created.

This explainable AI Model LIME can be used to text,image and tabular data and is adaptable with a wide range of classifiers.

- **Shapley Additive exPlanations**

Shapley Additive exPlanations is the abbreviation for SHAP. Shap is a framework for explainable AI that was derived from the game theory's Shapley values. Lundberg and Lee initially published this method in 2017.

The mean or standard marginal contribution of a feature value throughout all potential coalitions is known as the Shapley value. As it takes into account all conceivable permutations of input from the dataset when making predictions for an observation, SHAP offers a high consistency and local accuracy. This causes it to become computationally demanding. Different explainers that make some approximations are used to optimize the SHAP packages that are available (in R & Python). The SHAP results may be used to pinpoint exactly how each attribute contributed to a given prediction, making it simple for anybody to understand.

7.3. Other implementation details(if any)

- **Decision Tree Algorithm**
  - ➔ A supervised machine learning approach called a decision tree is used to solve classification and regression problems. Being a tree classifier, the internal nodes indicate the dataset's characteristics, while the branch symbolizes the decision-making process and the leaf node the outcome.
  - ➔ There are two different types of nodes in this algorithm: decision nodes and leaf nodes. While leaf nodes represent the results of those decisions and do not include any further branches, decision nodes are used to make decisions and are permitted to have numerous branches.
  - ➔ The features of the dataset determine the resolution/trials.This algorithm operates on the idea that a question is posed, and based on the boolean response (yes/no), it splits the tree further.

- **Random Forest Algorithm**

  - ➔ Random Forest is a prominent machine learning method that belongs to the supervised learning technique.Adopted for classification as well as regression issues in ML, it is established on the notion of ensemble learning which is the procedure of merging numerous classifiers to resolve a complicated difficulty and to upgrade the efficiency of the method.
  - ➔ As the name implies, Random Forest is a classifier that holds a number of decision trees on several subgroups of the provided dataset and then considers the average to enhance the prediction accuracy of the dataset.
  - ➔ Rather than depending on one decision tree this algorithm collects the output from each tree and predicts the concluding result.
  - ➔ Higher the number of trees , greater is the accuracy of the algorithm.

- **Gradient Boosted Decision Tree Algorithm**
  - ➔ Gradient Boosting Decision Tree is a technique enforced on top another machine learning algorithm.
  - ➔ Gradient boosted consists of two kinds of modules:

    - a weak machine learning algorithm that is generally a decision tree.

    - a strong machine learning algorithm that is composed of several weak models
  - ➔ In gradient boosting, a new weak model is trained at each step to anticipate the "error" of the existing strong model (which is called the pseudo response). Later, we'll go into more detail on "mistakes." Consider "error" for the time being to be the distinction between the prediction and a regressive label. Then, in order to lessen the mistake of the strong model, the weak model (i.e., the "error") is added to the strong model with a negative sign.

7.4. Discuss Data Set

We extracted a dataset for Cardiovascular Diseases with 12 features and 70000 entries of patient data from Kaggle. The concerned features were: Age, Height, Weight, Gender, Systolic blood pressure, Diastolic blood pressure, Cholesterol, Glucose, Smoking, Alcohol intake, Physical Activity, and Presence and absence of Cardiovascular disease.

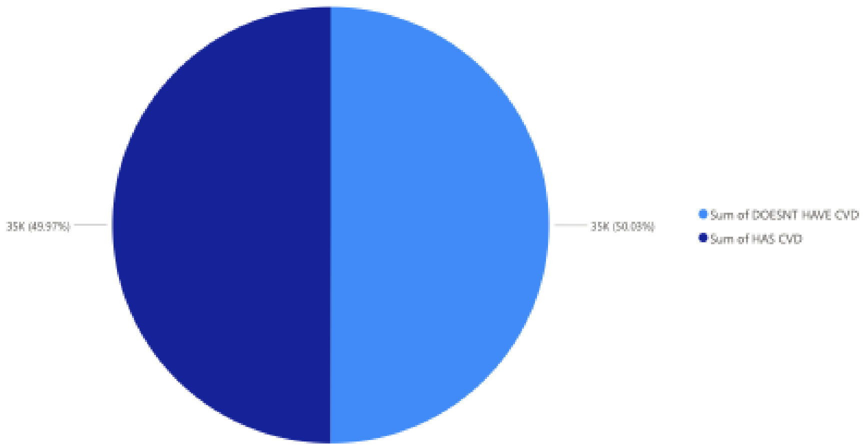


Fig. 7.4.1. Pie Chart for the count of ‘Has CVD’ and ‘Doesn’t have CVD’ classes

As you can see, we have a balanced dataset. Both the expected outputs, that is, “Has CVD” and “Doesn’t have CVD” have almost equal number of records in the dataset. Hence, moving forward, this dataset would be more feasible to use for our study.

After pre-processing this dataset and working on it for a while, we came to the conclusion that the main features that impacted our results were Cholesterol and Blood pressure of the patients.

Here’s a peek into those features:

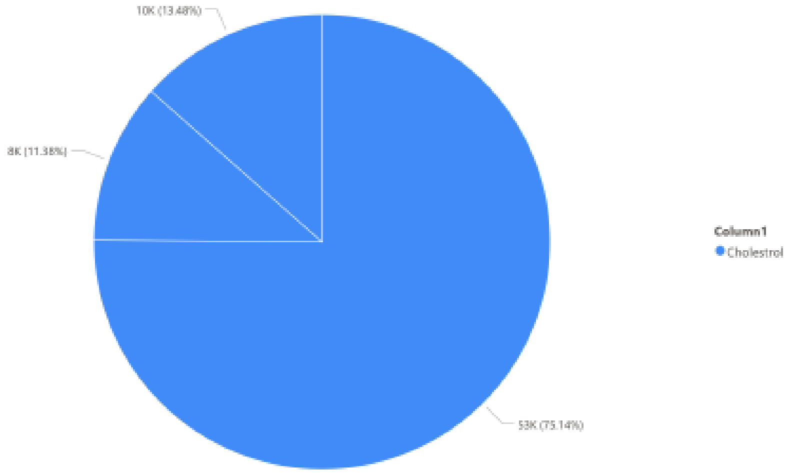


Fig. 7.4.2. Pie Chart showing record entries afflicted by Level 1, 2, and 3 Cholesterol

Above pie chart classifies the Cholesterol feature into three parts, i.e., at level 1, 2, and 3. According to the information available in the same context, we deduced that 53,000 patients fell into Level 1 category, while about 10,000 and 8,000 people were at Level 2 and 3 of Cholesterol rate.

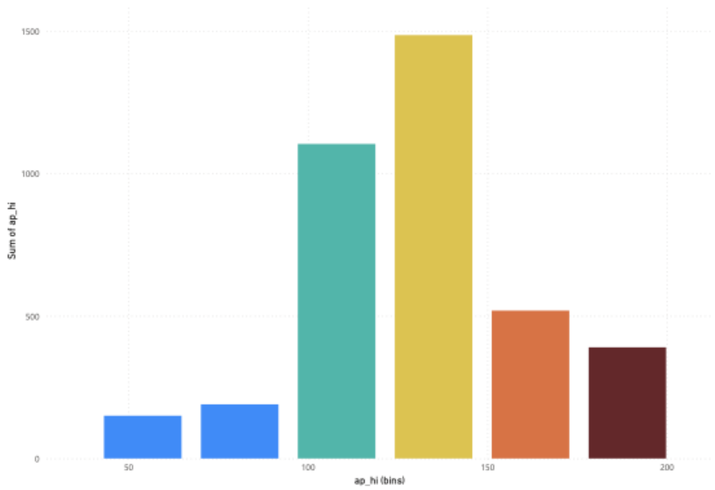


Fig. 7.4.3. Bar graph representing the status of blood pressure in concerned subjects

Next, we worked on the Blood pressure feature, creating a bar graph to understand the status of the patients. Here, while ‘Green’ bar represents average blood pressure, ‘Yellow’ points towards a slightly alleviated stage, and ‘Orange’ and ‘Maroon’ hints towards the dangerous and quite dire state of the feature. ‘Blue’ here signifies the number of patients with their blood pressure in control.

Chapter 8
Performance Evaluation and Testing

8.1. Test Cases

Project Name: Explainable AI in Healthcare

Test Case ID: TC_01

Test Priority (Low/Medium/High): High

Module Name: Data preprocessing

Test Title: Rectify and correct the data

Description: Test is data is ready to use

Test Designed by: Sakshi Kothari

Test Designed date: 14/11/2022

Test Executed by: Ritu Karnawat

Test Execution date: 16/11/2022

Pre-conditions: Data is downloaded and uploaded
Dependencies:

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	Null values are removed	if any value=0	No null values	No Null values	Pass	
2	Remove duplicate values	any value entered twice	No duplicate values	No duplicate values	Pass	
3	Drop ID	Table column= Id	No column as id	No column as id	Pass	
4	Adjust Age	Age is in months	Age converted into year	Age is in year	Pass	

Post-conditions:
Database is ready for machine learning processes

Table 8.1.1. TC_01: Rectify and correct the data

Project Name: Explainable AI in Healthcare

Test Case ID: TC_02

Test Priority (Low/Medium/High): Med

Module Name: Machine Learning Methods

Test Title: Random Forest

Description: Random Forest Classifier

Test Designed by: Veda Dhapke

Test Designed date: 14/11/2022

Test Executed by: Utkarsh Mukkawar

Test Execution date: 16/11/2022

Pre-conditions: Database is pre processed and ready to use
Dependencies:

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	Split Data	Dataset	Values splitted for training and testing purposes	Test Size=0.5	Pass	
2	Accuracy	Trained dataset	ac > 70%	ac=70.83	Pass	
3	Heatmap	Tested Dataset	Heatmap for confusion matrix	Heatmap of confusion matrix	Pass	
4	confusion matrix array	Tested dataset	array	([[5088, 1900], [2094, 4918]])	Pass	

Post-conditions:
Prediction needs to be explained through XAI methods

Table 8.1.2. TC_02: Random Forest

Project Name: Explainable AI in Healthcare

Test Case ID: TC_03

Test Designed by: Veda Dhapke

Test Priority (Low/Medium/High): Med

Test Designed date: 14/11/2022

Module Name: Machine Learning Methods

Test Executed by: Utkarsh Mukkawar

Test Title: Decision tree

Test Execution date: 16/11/2022

Description: Decision tree

Pre-conditions: Database is pre processed and ready to use
Dependencies:

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	Split Data	Dataset	Values splitted for training and testing purposes	Test Size=0.5	Pass	
2	Accuracy	Trained dataset	ac > 70%	ac=71.50	Pass	
3	Heatmap	Tested Dataset	Heatmap for confusion matrix	Heatmap of confusion matrix	Pass	
4	confusion matrix array	Tested dataset	array	([[13567,3863],[5622, 11948]])	Pass	

Post-conditions:
Prediction needs to be explained through XAI methods

Table 8.1.3. TC_03: Decision Tree

Project Name: Explainable AI in Healthcare

Test Case ID: TC_04

Test Priority (Low/Medium/High): Med

Module Name: Machine Learning Methods

Test Title: Gradient Boosted Tree

Description: Gradient Boosted Tree

Test Designed by: Veda Dhapke

Test Designed date: 14/11/2022

Test Executed by: Utkarsh Mukkawar

Test Execution date: 16/11/2022

Pre-conditions: Database is pre processed and ready to use

Dependencies:

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	Split Data	Dataset	Values splitted for training and testing purposes	Test Size=0.5	Pass	
2	Accuracy	Trained dataset	ac > 70%	ac=78.78	Pass	
3	Heatmap	Tested Dataset	Heatmap for confusion matrix	Heatmap of confusion matrix	Pass	
4	confusion matrix array	Tested dataset	array	([[14418,3012], [4353, 13217]])	Pass	

Post-conditions:

Prediction needs to be explained through XAI methods

Table 8.1.4. TC_04: Gradient Boosted Tree

Project Name: Explainable AI in Healthcare

Test Case ID: TC_05

Test Designed by: Sakshi Kothari

Test Priority (Low/Medium/High): High

Test Designed date: 14/11/2022

Module Name: XAI

Test Executed by: Ritu Karnawat

Test Title: LIME

Test Execution date: 16/11/2022

Description: using LIME explainer

Pre-conditions: all the machine learning predictions are completed

Dependencies:

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	LIME explainer applied to RF	Random Forest classifier model	Explainer of a given iloc	Explainer shown of iloc.669	Pass	
2	LIME explainer applied to DT	Decision tree classifier model	Explainer of a given iloc	Explainer shown of iloc.669	Pass	
3	LIME explainer applied to GBT	Gradient boosted tree classifier model	Explainer of a given iloc	Explainer shown of iloc.669	Pass	

Post-conditions:

Results sent for comparison

Table 8.1.5. TC_05: LIME

Project Name: Explainable AI in Healthcare

Test Case ID: TC_06

Test Designed by: Sakshi Kothari

Test Priority (Low/Medium/High): High

Test Designed date: 14/11/2022

Module Name: XAI

Test Executed by: Ritu Karnawat

Test Title: SHAP

Test Execution date: 16/11/2022

Description: using SHAP explainer

Pre-conditions: all the machine learning predictions are completed

Dependencies:

Step	Test Steps	Test Data	Expected Result	Actual Result	Status (Pass/Fail)	Notes
1	SHAPexplainer applied to RF	Random Forest classifier model	Explainer of a given iloc	Explainer shown of iloc.669	Pass	
2	SHAP explainer applied to DT	Decision tree classifier model	Explainer of a given iloc	only some features are shown	Fail	
3	SHAP explainer applied to GBT	Gradient boosted tree classifier model	Explainer of a given iloc	only some features are shown	Fail	

Post-conditions:

Results sent for comparison

Table 8.1.6. TC_06: SHAP

Chapter 9
Result and Analysis

9.1 Results of machine learning models:

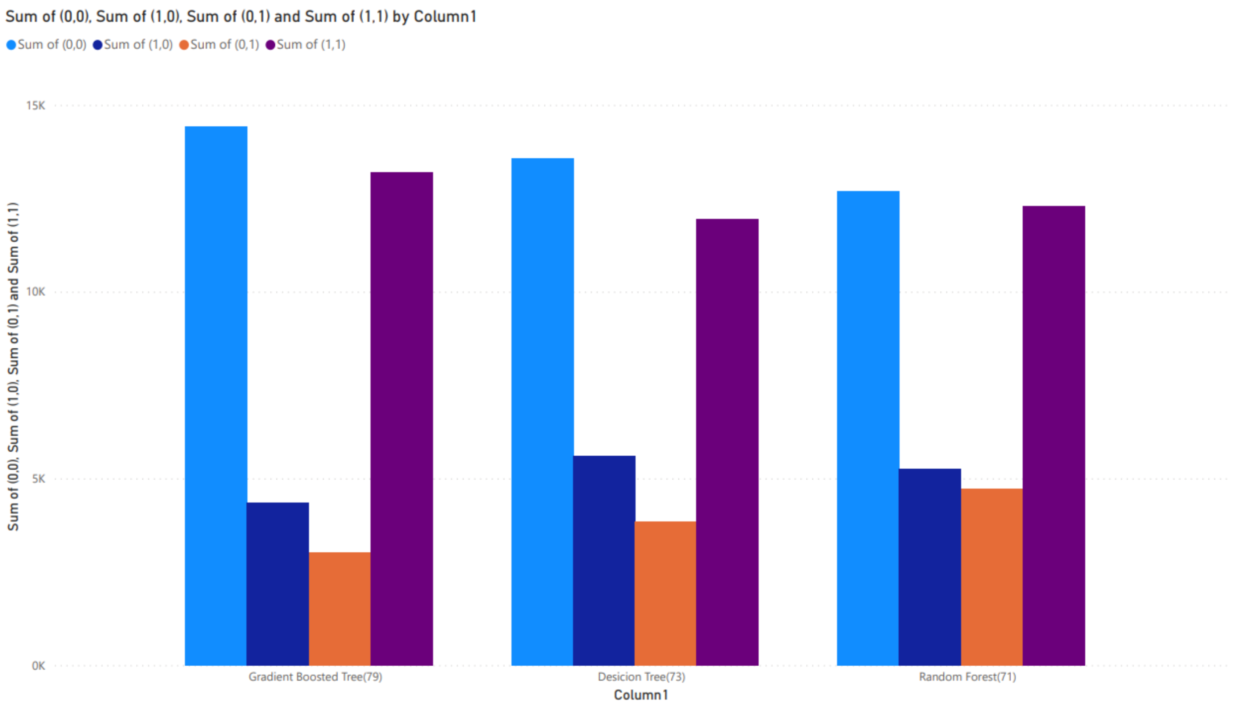


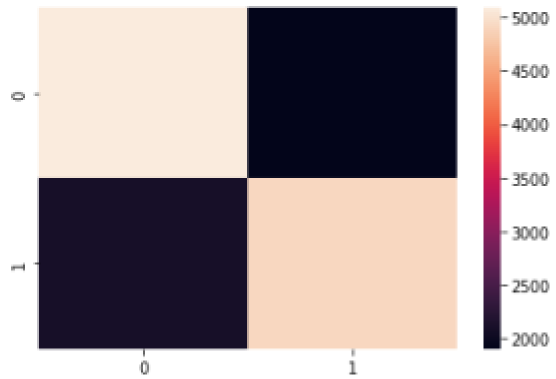
Fig. 9.1.1. Accuracy by machine learning models

In the above diagram the machine learning models are on the x axis and the sum of the values of (0,0),(0,1),(1,0),(1,1) are on the y axis which are sown in different colors form the figure we can conclude that the values of (0,0) and (1,1) are the highest which means the model has predicted the correct values most of the time which are true negative and true positive.

The ac tab below shows the accuracy of the respective models. The below heatmap describes different features which includes (0,0) which is true negative, (0,1) which is false positive, (1,0) which is false negative and (1,1) which is true positive. and the array below is the count of features listed above respectively.

- **Random Forest Classifier :**

```
ac

    0.7083928571428572

seaborn.heatmap(cm)

    <matplotlib.axes._subplots.AxesSubplot at 0x7fdd39724d50>
```

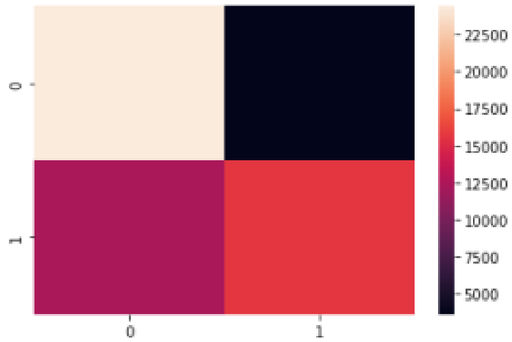



```
cm

    array([[5088, 1900],
           [2094, 4918]])
```

9.1.2. Random Forest classifier

- **Decision Tree Classifier :**

```
ac
    0.7150714285714286

seaborn.heatmap(cm)
    <matplotlib.axes._subplots.AxesSubplot at 0x7fdd395ba6d0>
```

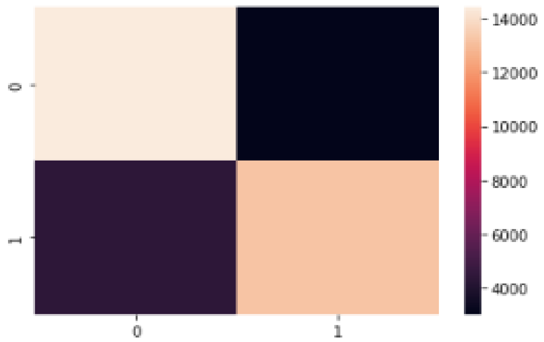



```
cm
    array([[13567,  3863],
           [ 5622, 11948]])
```

9.1.3. Decision Tree Classifier

- **Gradient Boosted Tree :**

```
ac
    0.7878035714285714

seaborn.heatmap(cm)
    <matplotlib.axes._subplots.AxesSubplot at 0x7fdd397f4b10>
```




```
cm
    array([[14418,  3012],
           [ 4353, 13217]])
```

9.1.4. Gradient Boosted Tree classifie

9.2 Results of Explainable AI Modes:

- **LIME(Local Interpretable Model-agnostic Explanations):**

```
test.iloc[[326]]
```

	age	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
32078	53.0	75.0	130	90	3	1	0	0	0	1

Fig. 9.2.1. Test subject

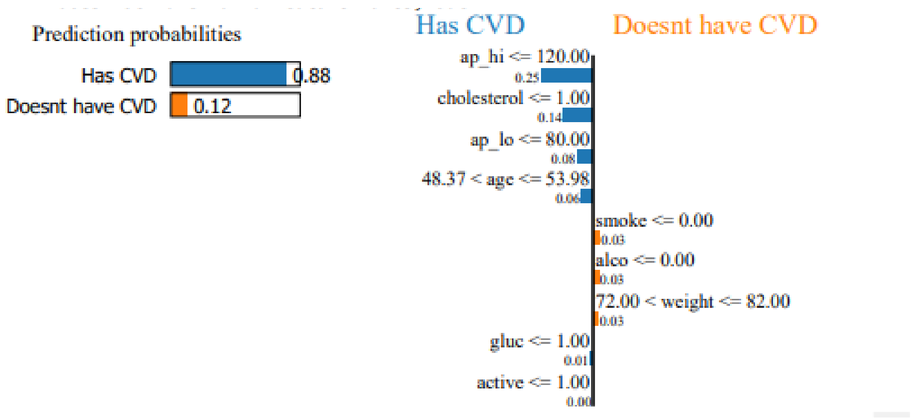


Fig. 9.2.2. LIME prediction for Random Forest

From the visualizations, we can conclude that the relatively higher price value(depicted by a bar on the left) of the CVD depicted by the given vector can be attributed to the following reasons:

1. The high value of cholesterol is the major factor for predicting the result and thus showing the same in the right side as the dominating feature for predicting the results.
2. Followed by the ap_hi and ap_lo are the high and low blood pressures of the patient and they are 0.14 and 0.05 amount 0.89 for the prediction Has CVD.
3. The features which are responsible for predicting the result Does Not Have CVD are age,gluc and active and the values they have contributed are 0.04,0.01 and 0 respectively.
4. The remaining features which contributed to the result Has CVD are smoke ,alco and weight.

From the above figure we can conclude that the lime has predicted the person has the chance of Having a CVD is 89 percent and by looking at the fig we can see that the value of cardio is 1 which indicates that the person having an active CVD

similarly the Results of other ML models are as below

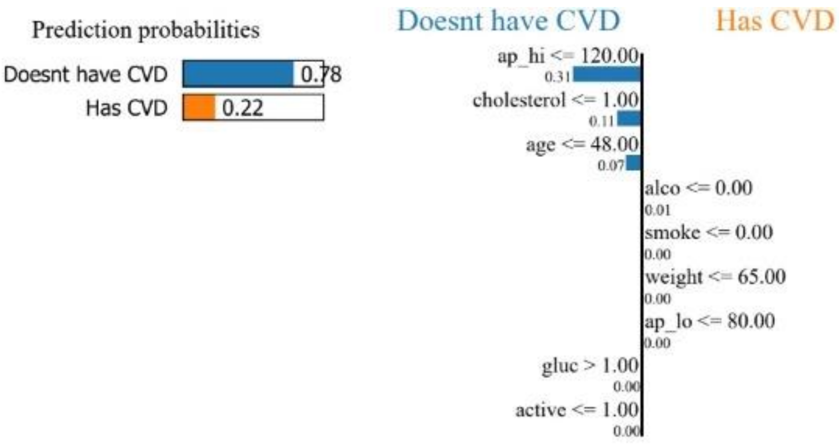


Fig. 9.2.3. LIME prediction for Decision Tree

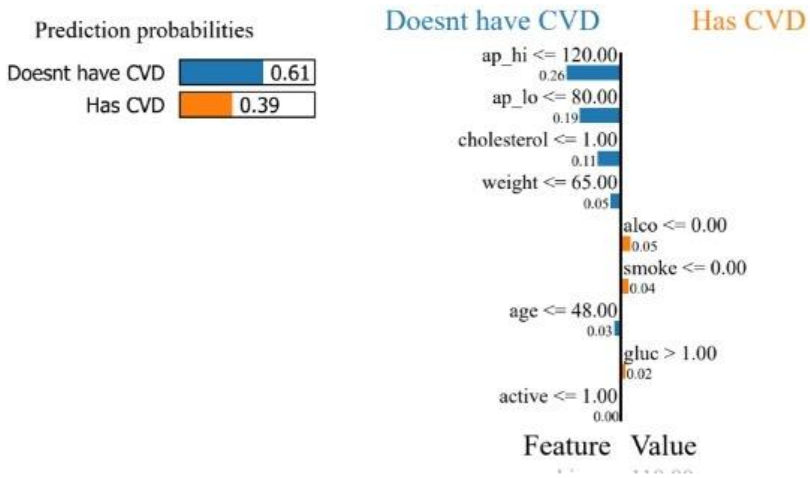


Fig. 9.2.4. LIME prediction for Gradient Boosted Tree

- **SHAP(Shapley Additive exPlanations):**

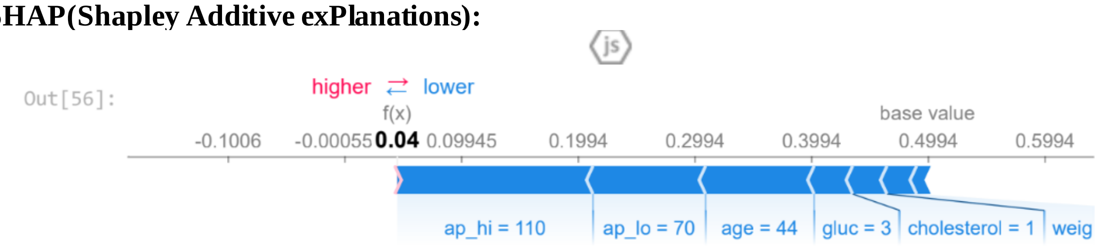
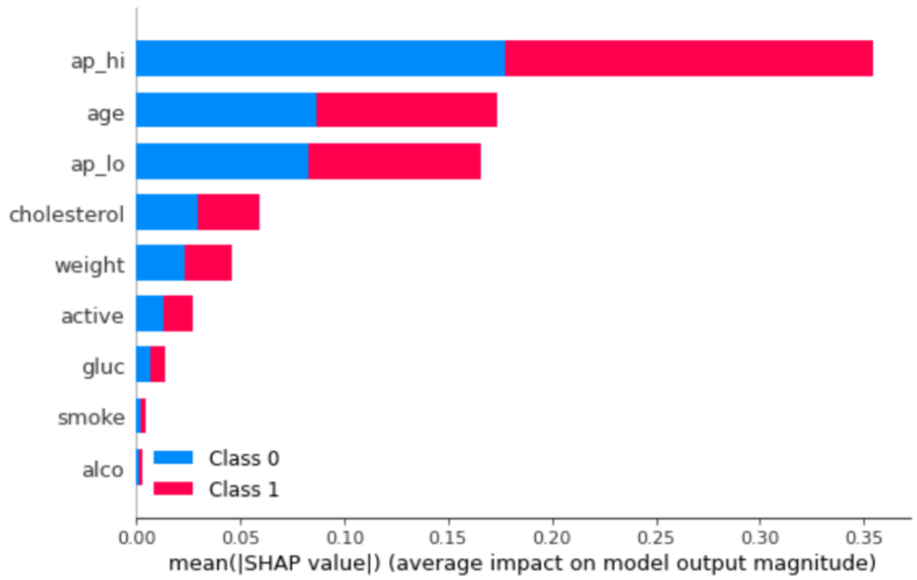


Fig. 9.2.5. SHAP prediction for Random Forest

From the visualizations, we can conclude that the relatively higher price value(depicted by a bar) of the Chances of survival (in terms of higher or lower) depicted by the given vector can be attributed to the following reasons:

1. The high value of ap_hi is indicating that the person has high blood pressure
2. Followed by the ap_lo which is low blood pressure.
3. The features which are responsible for predicting the result i.e. Lower chances of survival are age,gluc and cholesterol.
4. The remaining feature which contributed to the result is weight.



9.2.6. SHAP
feature impact

From the above figure we can see that the features are on the y axis and the mean of shap values is on x axis the bars which are in two colors i.e. blue and red are indicating class 0 and class 1 which are lower and higher chances of survival.

Applications

After a certain amount of work and accuracy increment of the model, we can have several applications of this research in the near future. The main application of our research is to include XAI in the healthcare sector so that the medical practitioners can easily figure out the patients who are prone to cardiovascular diseases and warn them in advance and take precautions to prevent it and deaths related to CVD.

Application of XAI will also develop trust of the medical practitioners on the prediction strategy of the machine learning algorithms and hence will ease the work of the professionals and also give assured diagnosis in advance based on several lifestyle habits of the patients..

Conclusion

We can learn more about how the AI model works by using explanation methods. These techniques still have a number of drawbacks, though. First, heatmaps produced using the explanatory methods used today depict "first-order" data, or the input features that have been determined to be important for the prediction. The relationship between these qualities, such as whether they are significant on their own or just when present together, is still unknown. Understanding these relationships is crucial for many applications, including healthcare, where these higher-order explanations may enable us to find multiple small factors for the disease, or groupings of cell areas that collaborate to exceed the risk, as opposed to just identifying significant individual voxels.

The explanations' low level of abstraction is another drawback. Without connecting these relevance values to more abstract ideas like the objects or the scene depicted in the image, heatmaps demonstrate which pixels are significant. To make sense of the explanations and comprehend the behavior of the model, humans must interpret them. This phase of interpretation may be challenging and inaccurate. Meta-explanations that gather data from these detailed heatmaps and describe the behavior of the model in a more abstract, approachable manner are preferred.

Future Scope

Our research limitations were that we weren't able to analyze test results received from SHAP because SHAP results were based on feature extraction and our model was only analyzing the prediction based on some features of the dataset, i.e., cholesterol and high blood pressure. Up until now, we worked on only three machine learning techniques, but moving forward, we could increase our scope to several deep learning models, alongside different explainers. Not to mention, we could also work on different categories in cardiovascular diseases based on features like smoking, state of your arteries, blood type, atrial blockage, etc. Therefore , our future scope for this research includes working on SHAP such that it considers the other factors as well for the explanation(like smoking, alcohol,etc.)

References

[1] Manisha Chandak, Sheetal Girase, Introducing hybrid technique for optimization of book recommender system, International Conference on Advanced Computing Technologies and Applications (ICACTA-2015), Procedia Computer Science page nos- 23 – 31, ELSEVIER

[2] "Smart healthcare: making medical care more intelligent," Global Health Journal, vol. 3, no. 3, pp. 62–65, 2019.

[3] S. Khedkar, V. Subramanian, G. Shinde, and P. Gandhi, "Explainable AI in Healthcare," SSRN Electronic Journal, 2019.

[4] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," no. Ml, pp. 1–13, 2017.

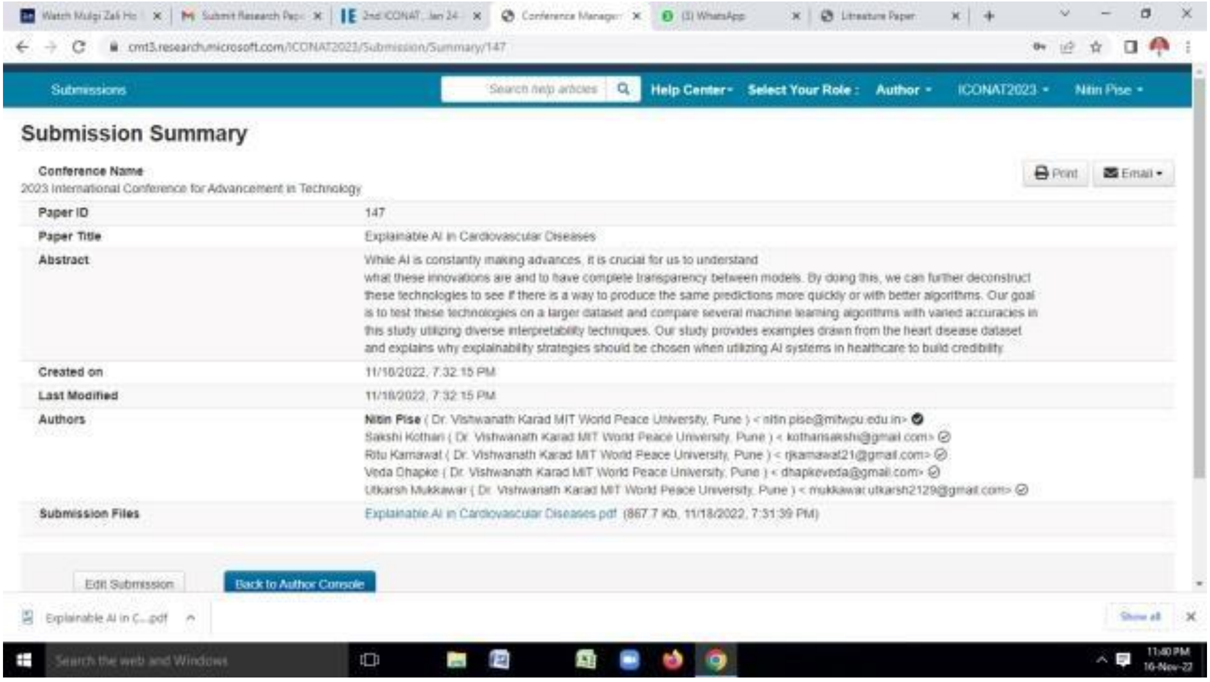
[5]Westerlund, A.M.; Hawe, J.S.; Heinig, M.; Schunkert, H. Risk Prediction of Cardiovascular Events by Exploration of Molecular Data with Explainable Artificial Intelligence. Int. J. Mol. Sci. 2021, 22, 10291. https://doi.org/10.3390/ ijms221910291

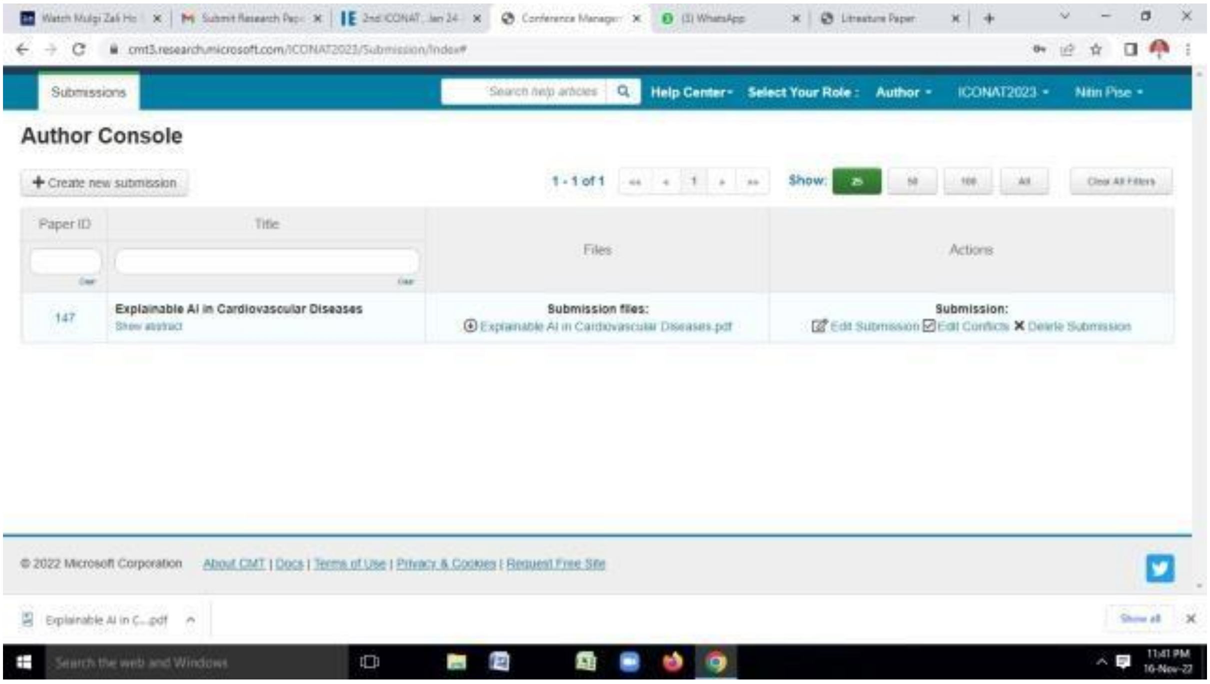
[6]Pawar, Urja & O'Shea, Donna & Rea, Susan & O'Reilly, Ruairi. (2020). Explainable AI in Healthcare. 10.1109/CyberSA49311.2020.9139655.

[7]Cinà, Giovanni & Röber, Tabea & Goedhart, Rob & Birbil, Ilker. (2022). Why we do need Explainable AI for Healthcare. 10.48550/arXiv.2206.15363.

[8]Di Martino, F., Delmastro, F. Explainable AI for clinical and remote health applications: a survey on tabular and time series data. *Artif Intell Rev* (2022). https://doi.org/10.1007/s10462-022-10304-3

Conference Name: 2023 International Conference for Advancement in Technology.





Explainable AI in Cardiovascular Diseases

Prof. Nitin.Pise
School of Computer Engineering and technology
MIT -WPU,
Pune, Maharashtra, India
nitin.pise@mitwpu.edu.in

Sakshi. Kothari
School of Computer Engineering and technology.
MIT -WPU,
Pune, Maharashtra, India
kotharisakshi103@gmail.com

Ritu. Karnawat
School of Computer Engineering and technology
MIT -WPU,
Pune, Maharashtra, India
rjkarnawat21@gmail.com

Veda. Dhapke
School of Computer Engineering and technology
MIT -WPU,
Pune, Maharashtra, India
dhapkeveda@gmail.com

Utkarsh. Mukkawar
School of Computer Engineering and technology
MIT -WPU,
Pune, Maharashtra, India
mukkawar.utkarsh2129@gmail.com

Abstract- While AI is constantly making advances, it is crucial for us to understand what these innovations are and to have complete transparency between models. By doing this, we can further deconstruct these technologies to see if there is a way to produce the same predictions more quickly or with better algorithms.

Our goal is to test these technologies on a larger dataset and compare several machine learning algorithms with varied accuracies in this study utilising diverse interpretability techniques.Our study provides examples drawn from the heart disease dataset and explains why explainability strategies should be chosen when utilising AI systems in healthcare to build credibility.

Keywords: Explainable AI, Healthcare, Heart disease, Programming frameworks, LIME, SHAP, Machine Learning Techniques.

I. Introduction

According to estimates, 17.9 million people worldwide die from cardiovascular disorders each year (Reference: WHO statistics). Heart attacks and strokes are to blame for four out of every five CVD fatalities. Heart disease and stroke are primarily caused by bad food, lack of exercise, smoking, alcohol consumption, and obesity, among other things.

For people and medical professionals to make decisions and be sure that the results are reliable, explanation of the fundamental reasoning is vital for CVDs and other healthcare applications. Lack of explainability in machine learning limits the expansive application of AI.[1] If AI is unable to explain how it predicts outcomes in the healthcare industry, the danger of making a bad choice may outweigh its benefits of precision and speed, which would restrict its use. Standard tools must be created in order to tackle these problems and use AI to predict CVDs in advance and take preventative measures. Explainable AI is one such tool (XAI). It is possible to use XAI to explain how and why a person is predicted to develop cardiac disease in the near future. As a result, medical professionals would have more confidence in AI, which would allow them to treat patients earlier and lower the death rates from CVDs.

Health information collected at a user level can also be shared with clinicians for further diagnosis [1] and together with AI can be used in health screening, early diagnosis of diseases, and treatment plan selection [2]. In the healthcare domain, the ethical issue of transparency associated with AI and the lack of trust in the black-box operation of AI systems creates the need for AI models that can be explained [3]. Further XAI integration into intelligent healthcare systems is possible for more disorders including cancer, neurology, etc. These technologies can be used to diagnose serious illnesses and choose the best course of therapy. Our research aims to describe the many XAI methods used on the Kaggle dataset for cardiovascular disorders, including SHapley and LIME.

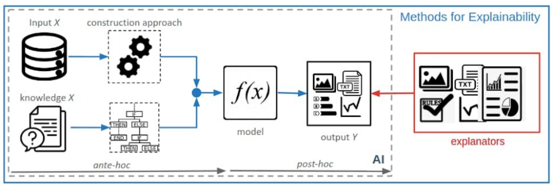


Fig.1.1: Brief Diagram of Explainable AI

II. Literature Review

[4] The themes of traditional risk prediction models, artificial intelligence (AI) used to recognise CVD in various types of imaging data, or AI for the purposes of drug discovery and electrocardiogram-based prediction have all received a lot of attention recently. Reviews that take into account the molecular and genetic elements of CVDs have also been published recently, with the majority of them concentrating on detailing the intricate processes that underlie disease. These reviews take incident CVD into account as opposed to recurring occurrences. Here, however, we concentrate on recent developments in risk prediction of recurrent cardiovascular events based on clinical and molecular data and highlight the application of AI as a tool for (1) forecasting risk and (2) developing individualised treatment plans. Because performance measures are frequently used to objectively evaluate prediction models,Finally, we discuss how explainable AI (XAI) may be a key component of future transparent and dependable clinical decision systems that are based on both clinical and molecular data. These issues in risk prediction can now be handled with AI.

[5]One of the most common model-agnostic approaches is called Local Interpretable Model-Agnostic Explanation (LIME), which is a framework for explaining predictions by quantifying the contributions of all the factors involved in calculating predictions. Researchers in used LIME to describe how Recurrent Neural Networks (RNNs) anticipate heart failure, and their explanations helped pinpoint the most common illnesses that raise a person's risk of developing heart failure, such as diabetes, anaemia, and renal failure. Anchors and Shapley values are two further model-independent XAI

techniques that have been created and are used in the healthcare sector.

[6]We have looked at some of the main arguments put forth in the literature against the pervasive application of XAI methods in medical settings. Some of the concerns and arguments put up in response to them can actually be reframed to apply to applications outside of the healthcare industry. However, we should be aware of how the discussion is affected by the increased stakes involved in safety-critical AI applications in the healthcare industry. It is important to note that downgrading explainable medical AI's usefulness would mean suppressing efforts to obtain human oversight over a cutting-edge technology that is still in its infancy but has the potential to be quickly and widely implemented, potentially affecting a large portion of the population.XAI is not a panacea for all of AI's issues, and it is especially not a replacement for meticulous model performance analysis. However, there are a number of situations where XAI technologies can be helpful. More black boxes won't stop the loss of human control in a rapidly digitising society since human monitoring of machines is still a crucial principle. Enhancing human capabilities has been claimed to be better to automation since the latter causes deskilling, detachment, and increases in unemployment.The prospect of a human-machine synergy, or a means for people to interact with AI, is assumed by augmentation. The European Union and other public organisations are debating rules for AI goods. Although the XAI field may not yet be mature enough for XAI techniques to be included into rules as hard requirements, this does not mean that XAI approaches should be ignored or put on the back burner. In order to exert an additional layer of control over AI products, the aforementioned strategies can be used; however, as with any technology, users should be aware of any potential failure modes. In terms of rules,Allowing the usage of XAI techniques while requesting an explanation as to why a particular technique was selected could be useful.

III. Dataset description

There are 3 types of input features:

Features:

Objective Feature

1. Age | int (days)
2. Height |int (cm) |
3. Weight | float (kg) |
4. Gender | categorical code |

Examination Feature

5. Systolic blood pressure | int |
6. Diastolic blood pressure | int |
7. Cholesterol | 1: normal, 2: above normal, 3: well above normal |
8. Glucose | 1: normal, 2: above normal, 3: well above normal |

Subjective Feature

9. Smoking | binary |
10. Alcohol intake alco | binary |
11. Physical activity | active | binary |

12. Presence or absence of cardiovascular disease | cardio | binary |

All of the dataset values were collected at the moment of medical examination.

IV. Methods

This paper provides an overview of various approaches to explainable AI, starting with methods that don't depend on specific models and instead use a straightforward surrogate function to explain the predictions.

4.1Local Interpretable Model-agnostic Explanations (LIME)

LIME which is Local Interpretable Model-agnostic Explanations is an explainable AI technique which helps in throwing light on a machine learning model and making every prediction's specific inference understandable.Due to the fact that it describes the classifier for a specific single instance, this technique is appropriate for local explanations.

LIME modifies the input data to produce or generate a sequence of fictional data with just a minimal amount of the original properties retained. As a result, multiple replicas of original text are created.

This explainable AI Model LIME can be used to text,image and tabular data and is, adaptable with a wide range of classifiers.

4.2.Shapley Additive exPlanations

Shapley Additive exPlanations is the abbreviation for SHAP. Shap is a framework for explainable AI that was derived from the game theory's Shapley values. Lundberg and Lee initially published this method in 2017.

The mean or standard marginal contribution of a feature value throughout all potential coalitions is known as the Shapley value.

As it takes into account all conceivable permutations of input from the dataset when making predictions for an observation, SHAP offers a high consistency and local accuracy. This causes it to become computationally demanding. Different explainers that make some approximations are used to optimise the SHAP packages that are available (in R & Python).

The SHAP results may be used to pinpoint exactly how each attribute contributed to a given prediction, making it simple for anybody to understand.

Create your own databases using the Analyttica TreasureHunt® LEAPS platform, run the models, and be able to explain the results.

V. Proposed Methodology

In this study, we classified our dataset using three machine learning techniques, i.e., Random Forest, Gradient Boosted Tree, and Decision Tree. Moving forward, we elaborated the given methods using Explainable AI technique- LIME, which builds a local interpretable model with the intent of calculating feature values.

Provided below is the LIME explainer code we have implemented in our analysis:

```
predict_fn_rf = lambda x: random_forest.predict_proba(x).astype(float)
X = X_train.values
```

explainer=lime.lime_tabular.LimeTabularExplainer(X,feature_names=X_train.columns,class_names=['Doesn't have CVD','Has CVD'],kernel_width=5)

Herein, we have initially calculated the prediction probability of our data using the Random Forest predict_proba method. This method estimates class probabilities for the input x. These probabilities are calculated as average predicted class probabilities of the trees in the forest.

We have stored the trained values in a variable and then written the explained code which uses the lime_tabular.LimeTabularExplainer method which takes the parameters as x_train.columns and gives class names as "Doesn't have CVD" and "Has CVD".

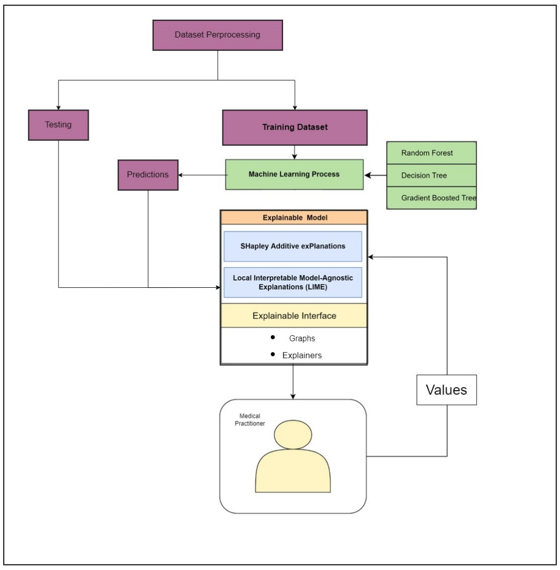


Fig.5.1: System Architecture Diagram

VI. Results

We selected a particular location[669] in the dataset and then explained the three Machine Learning techniques using LIME explainer code.

	age	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
42091	44.0	61.0	110	70	1	3	0	0	1	0

Fig.6.1: Data at location 669

The result is given probable occurrences of the two classes, which are, "Doesn't have CVD" and "Has CVD".

For instance, in case of Gradient Boosted Tree, the prediction probability for "Doesn't have CVD" is 0.61, whereas for "Has CVD" is 0.39. The result is primarily affected by the factors that are highlighted in 'Orange' like alco, smoke, and gluc in this specific location. Evidently, both alco and smoke has value 0, which directs towards the patient's possibility of acquiring CVD as 'low'.

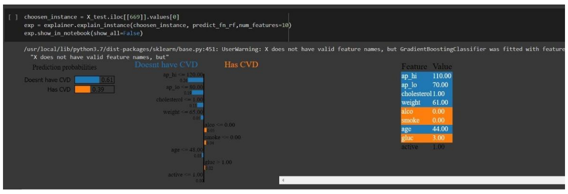


Fig.6.2: LIME result for gradient boosted tree.

Similarly we have applied LIME on the other two ML techniques- Decision Tree and Random Forest.

-Decision Tree: Prediction probability for "Doesn't have CVD" is 0.78 and for "Has CVD" is 0.22.

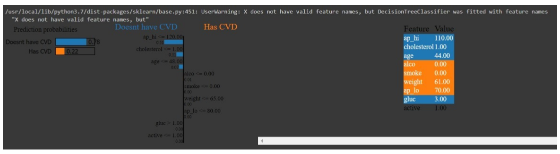


Fig.6.3:LIME result for Decision Tree

-Random Forest: Prediction probability for "Doesn't have CVD" is 0.92 and for "Has CVD" is 0.09.

Fig.6.4: LIME result for Random Forest

VII. Conclusion

We can learn more about how the AI model works by using explanation methods. These techniques still have a number of drawbacks, though. First, heatmaps produced using the explanatory methods used today depict "first-order" data, or the input features that have been determined to be important for the prediction. The relationship between these qualities, such as whether they are significant on their own or just when present together, is still unknown. Understanding these relationships is crucial for many applications, including healthcare, where these higher-order explanations may enable us to find multiple small factors for the disease, or groupings of cell areas that collaborate to exceed the risk, as opposed to just identifying significant individual voxels.

The explanations' low level of abstraction is another drawback. Without connecting these relevance values to more abstract ideas like the objects or the scene depicted in the image, heatmaps demonstrate which pixels are significant. To make sense of the explanations and comprehend the behaviour of the model, humans must interpret them. This phase of interpretation may be challenging and inaccurate. Meta-explanations that gather data from these detailed heatmaps and describe the behaviour of the model in a more abstract, approachable manner are preferred.

VIII. References

[1]Di Martino, F., Delmastro, F. Explainable AI for clinical and remote health applications: a survey on tabular and time series data. *Artif Intell Rev* (2022). https://doi.org/10.1007/s10462-022-10304-3

[2] “Smart healthcare: making medical care more intelligent,” Global Health Journal, vol. 3, no. 3, pp. 62–65, 2019.

[3] S. Khedkar, V. Subramanian, G. Shinde, and P. Gandhi, “Explainable AI in Healthcare,” SSRN Electronic Journal, 2019.

[4] F. Doshi-Velez and B. Kim, “Towards A Rigorous Science of Interpretable Machine Learning,” no. Ml, pp. 1–13, 2017.

[5]Westerlund, A.M.; Hawe, J.S.; Heinig, M.; Schunkert, H. Risk Prediction of Cardiovascular
Events by Exploration of Molecular Data with Explainable Artificial Intelligence. Int. J. Mol. Sci. 2021, 22, 10291. https://doi.org/10.3390/ ijms221910291

[6]Pawar, Urja & O'Shea, Donna & Rea, Susan & O'Reilly, Ruairi. (2020). Explainable AI in Healthcare. 10.1109/CyberSA49311.2020.9139655.

[7]Cinà, Giovanni & Röber, Tabea & Goedhart, Rob & Birbil, Ilker. (2022). Why we do need Explainable AI for Healthcare. 10.48550/arXiv.2206.15363.

PAPER NAME

Litreature Paper_A115.11.22.pdf

AUTHOR

U M

WORD COUNT

2284 Words

CHARACTER COUNT

12590 Characters

PAGE COUNT

6 Pages

FILE SIZE

674.6KB

SUBMISSION DATE

Nov 15, 2022 4:28 PM GMT+5:30

REPORT DATE

Nov 15, 2022 4:29 PM GMT+5:30

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 10% Internet database
- 12% Publications database
- Crossref database
- Crossref Posted Content database
- 8% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 8 words)