Enhancing Heart Disease Prediction with Explainable AI (XAI)

MS Thesis

By

Mubashir Iqbal

CIIT/SP22-RCS-002/WAH

COMSATS University Islamabad

Wah Campus - Pakistan

Fall, 2023

Enhancing Heart Disease Prediction with Explainable AI (XAI)

A Thesis submitted to

COMSATS University Islamabad, Wah Campus

In partial fulfillment

of the requirement for the degree of

MS in Computer Science

By

Mubashir Iqbal

CIIT/SP22-RCS-002/WAH

Department of Computer Science

Faculty of Information Science and Technology

COMSATS University Islamabad

Wah Campus - Pakistan

Fall, 2023

Enhancing Heart Disease Prediction with Explainable AI (XAI)

This thesis is submitted to the Department of Computer Science as a partial fulfillment of the requirement for the award of a Degree of MS in Computer Science.

Name	Registration Number
Mubashir Iqbal	CIIT/SP22-RCS-002/WAH

Supervisory Committee

Supervisor
Dr. Kashif Ayyub
Assistant Professor
Department of Computer Science
COMSATS University Islamabad
Wah Campus

Member
Dr. Muhammad Wasif Nisar
Professor
Department of Computer Science
COMSATS University Islamabad
Wah Campus

Member
Dr. Ehsan Ullah Munir
Professor
Department of Computer Science
COMSATS University Islamabad
Wah Campus

Member
Dr. Tassawar Iqbal
Associate Professor
Department of Computer Science
COMSATS University Islamabad
Wah Campus

Certificate of Approval

This is to certify that the research work presented in this thesis, entitled "Enhancing Heart Disease Prediction with Explainable AI (XAI)" was conducted by Mubashir Iqbal CIIT/SP22-RCS-002, under the supervision of Assistant Professor Dr. Kashif Ayyub. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the Department of Computer Science, COMSATS University Islamabad, Wah Cantt, in partial fulfillment of the requirements for the degree of MS in the field of Computer Science.

Student Name: Mubashir Iqbal Signature: _____

Examinations Committee:

<<External Examiner 1: Name>>
(Designation & Office Address)
……

<<External Examiner 2: Name>>
(Designation & Office Address)
……

Dr. Kashif Ayyub
Supervisor
Department of Computer Science
COMSATS University Islamabad (CUI)
Wah Cantt

Prof. Sheraz Anjum
Head Department of Computer Science
COMSATS University Islamabad (CUI)
Wah Cantt

Prof. Dr Ehsan Ullah
Chairperson
Computer Science
COMSATS University Islamabad (CUI)

Prof. Dr. Zulfiqar Habib
Dean
Information Science and Technology
COMSATS University Islamabad (CUI)

Author's Declaration

I, Mubashir Iqbal, CIIT/SP22-RCS-002/WAH, hereby state that my MS thesis titled "Enhancing Heart Disease Prediction with Explainable AI (XAI)" is my own work and has not been submitted previously by me for taking any degree from this University i.e. COMSATS University Islamabad or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after I graduate, the University has the right to withdraw my MS degree.

Dated: _____

Mubashir Iqbal

CIIT/SP22-RCS-002/WAH

Plagiarism Undertaking

I solemnly declare that the research work presented in the thesis titled Enhancing Heart Disease Prediction with Explainable AI (XAI)" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of HEC and COMSATS University Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake if I am found guilty of any formal plagiarism in the above titled thesis even after award of Ph.D. Degree, the University reserves the right to withdraw/revoke my Ph.D. degree and that HEC and the university has the right to publish my name on the HEC/university website on which names of students are placed who submitted plagiarized thesis.

Dated: _____

Mubashir Iqbal

CIIT/SP22-RCS-002/WAH

Certificate

It is certified that Mubashir Iqbal, CIIT/SP22-RCS-002/WAH has carried out all the work related to this thesis under my supervision at the Department of Computer Science, COMSATS University Islamabad, Wah Campus and the work fulfills the requirements for the award of the degree of MS in Computer Science.

Date: _____

Supervisor:

_____

Dr. Kashif Ayyub
Assistant Professor
Department of Computer Science
COMSATS University Islamabad
Wah Campus

DEDICATION

To ALLAH Almighty and His Last Beloved Prophet Muhammad (P.B.U.H)

&

My Parents, My Daughters, and all Teachers

ACKNOWLEDGEMENTS

I complete this research work with the help of **ALLAH Almighty** who always blesses me, forgives me, and guides me toward the rightest path of Jannah. Allah is the one who always blesses me with His endless treasures and limitless Mercy. Indeed, I could have done nothing without His permission and guidance. May Allah love us all a lot and give us the ultimate reward in the form of Jannah on Judgment Day.

I would like to acknowledge my supervisor Dr. Kashif Ayyub for all his diligence, guidance, and supervision that enabled me to complete this research work. Finally, I am profusely thankful to my parents who brought me to this stage of life, and their guidance that is an encouragement for me in every step of my life. Also, I am very grateful to my friends Hassan Shah, Hassan Sardar, and Asad Mashood for their continuous support, motivation, and encouragement throughout my MS journey.

Mubashir Iqbal
CIIT/SP22-RCS-002/WAH

ABSTRACT

Enhancing Heart Disease Prediction with Explainable AI (XAI)

By

Mubashir Iqbal

Predicting cardiovascular health stands as a critical imperative in modern healthcare, demanding the deployment of sophisticated and potent predictive models. The study seeks a Multi-Layer Perceptron (MLP) architecture network model for the predicting of heart disease. The dataset consists of 51 feature columns with a large records base of 37079 depicting different patient profiles. This study shows that the proposed architecture of the MLP model effectively captures deep patterns within the data. The research journey begins with specific care in preprocessing. It was not merely about compiling data: filling in the blanks, deleting duplicates, and finally using the Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic (ADASYN) to correct the original class imbalance problems for oversampling the minority samples. After describing the proposed architecture, the model has undergone extensive training and evaluation, resulting in notable results. The model has achieved a test accuracy of 97.10%. The model has once again proven its capability of identifying a positive case with sensitivity and precision values of 97.85% and 96.40% respectively while keeping away false positives. In a process where abstract numbers take shape, the explainability of the model provides a whole new dimension of transparency in the decision-making process.

The study shows the top features using the SHAP values and plots the summary (swarm) and bar plots according to their weightage. Through the SHAP technique, the proposed model gives transparency, and traceability and gets the medical staff's trust and credibility in decision-making. XAI algorithms like (SHAP, LIME, FairML, etc) are redefining the AI model's architectures, implementation, scope, deployment criteria, transparency, and trustworthiness in their decision or outcome.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AAMI	Association for Advancement of Medical Instrumentation
ANNs	Artificial Neural Network
AUC	Area Under the Curve
AUPRC	Area under the precision-recall curve
AUROC	Area under the receiver operating characteristic curve
BERHT	Bidirectional Encoder Representations from Transformers
BL	Binary Logistic
BPNN	Back Propagation Neural Network
CAD	Coronary artery disease
CAS	Carotid artery stenting
CDSS	Clinical Decision Support Systems
CHD	Coronary heart disease
CNN	Convolutional Neural Networks
CPRD	Clinical Practice Research Datalink
CRT	Resynchronization therapy
CVD	Cardiovascular disease
CXplain	Causal Explanations
DARPA	Defense Advanced Research Projects Agency
DCNN	Deep Convolutional Neural Network
DeepLIFT	Deep Learning Important FeaTures
DL	Deep Learning
DT	Decision Tree
ECG	Electrocardiography
EGM	Electrogram
ELU	Exponential Linear Unit
FC	Feature contributions
FL	Fuzzy Logic
FRC	Feature Ranking Cost
FRS	Framingham Risk Scores
FS	Feature subset
FW	Feature weights
GA	Genetic Algorithm
GD	Gradient Descent
HER	Electronic health records
HOBDBNN	Higher order Boltzmann deep belief neural network
HRFLM	Hybrid random forest with a linear model
HSP	Heart Surface Potential
IHD	Ischaemic heart disease
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IOT	Internet of things

KNN	K-nearest neighbor
LIME	Local Interpretable Model-agnostic Explanations
LR, LRC	Logistic Regression Classifier
LSD	Logarithmic standard deviation
LSTM	Long Short-Term Memory
MACE	Major adverse cardiovascular events
MCDM	Multi-Criteria Decision Making
MDCNN	Modified Deep Convolutional Neural Network
ML	Machine learning
MMRE	Mean magnitude of the relative error
MOEA	Multi-Objective Evolutionary Algorithm
NHANES	National Health and Nutritional Examination Survey
NLP	Natural language processing
PDP	Partial Dependence Plot
QoS	Quality of Service
ReLU	Rectified linear unit
RF, RFC	Random Forest Classifier
RNN	Recurrent Neural Networks
SEE	Software Effort Estimation
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-sampling Technique
SNP	Single nucleotide polymorphism
SVM	Support Vector Machine
TC	Traffic Classification
TNHIRD	Taiwan National Health Insurance Research Database
WHO	World Health Organization
XAI	eXplainable Artificial Intelligence
XGB	XGBoost

Chapter 1: Introduction

Coronary heart disease stands as one of the most fatal ailments globally, claiming the lives of approximately a third of the global population. It manifests when the heart faces challenges in efficiently pumping blood, often triggered by various factors including arterial sclerosis, hypertension, and hyperlipidemia. If cardiac infarction is found and treated early, a person's chances of survival are much better. That's why researchers are developing ways to predict who is at high risk for heart disease. These prediction models could be used to identify people who need to take steps to prevent heart disease or manage it early on [1].

The World Health Organization (WHO) reports that cardiovascular disease (CVD) is Pakistan's top cause of mortality, accounting for almost 30% of all fatalities annually. A 2022 study [2] published in the journal Heart found that the prevalence of Ischaemic Heart Disease (IHD) in Pakistan is 17.0%. IHD is a kind of CVD that happens when the heart's blood flow is restricted or obstructed. This can lead to a heart attack. The same study found that the following risk factors mentioned in Table 1.1, for CVD are highly prevalent in Pakistan.

Table 1.1: Risk Factors for CVD in Pakistan

S/N	Risk Factors	Percentage
1	Hypertension (high blood pressure)	40.1
2	Diabetes	15.8
3	Overweight/obesity	68.8
4	Tobacco	13.6

A healthcare practitioner will evaluate patients and delve into their individual and familial medical backgrounds. Diagnosing heart disease involves a range of tests. Alongside chest X-rays and blood examinations, additional diagnostic techniques for heart disease encompass:

- **Electrocardiogram (ECG)** The electrical signals within the heart can be captured through a straightforward and painless examination known as an Electrocardiogram (ECG). This diagnostic tool possesses the capability to identify irregular heartbeats.
- **Holter monitoring** An electrocardiogram (ECG) equipment that is portable and used to record the heart's activity while performing daily tasks is called a Holter monitor. It is worn for one or more days. An abnormal heartbeat that is missed by a routine ECG examination can be identified with this test.
- **Echocardiogram** Sound waves are used in this non-invasive examination to provide detailed images of the beating heart. It depicts how blood passes via the heart's valves. The presence of narrowing or leakage in a valve can be detected with an echocardiography.
- **Exercise tests or stress tests** Often, this involves walking on a treadmill or cycling in a stationary position while the heart is monitored. Exercise testing can reveal whether a person has heart disease symptoms and how their heart responds to physical strain.
- **Cardiac catheterization** This test can identify blockages in the heart arteries. A catheter is a long, thin, flexible tube that is inserted into a blood vessel, usually in the groin or wrist, and guided to the heart. The

arteries of the heart can be dyed thanks to the catheter. The dye increases the visibility of the arteries on X-ray images during the examination.

- **Heart CT scan.** A heart CT scan is performed while you are lying on a table with a doughnut-shaped scanner. The device rotates an X-ray tube around the human body to obtain images of the heart and chest.
- **Heart Magnetic Resonance Imaging (MRI)** A cardiac MRI creates extremely detailed images of the heart using a magnetic field and computer-generated radio waves.

The high burden of heart disease in Pakistan is probably a result of these risk factors' high prevalence. Recognizing the need to address heart disease, the Pakistani government has created a number of measures to reduce the disease's burden. These programs involve encouraging healthy lifestyles, increasing access to early detection and treatment services, and increasing public knowledge of heart disease and its risk factors. Still, more work has to be done to overcome this health problem. Investing in research to create more effective methods for heart disease diagnosis, treatment, and prevention is part of this.

1.1 Background

Cardiac catheterization is a minimally invasive technique utilizing a cylindrical, flexible tube (catheter) to analyze the heart's electrical activity. The catheter is inserted into a vein in the leg and paced up to the heart, exactly at the ventricles or atria. Electrodes at the tip of the catheter are utilized to measure the electrical signals generated from the cardiac muscle. Electrograms (EGMs) are these measurements. The signal represents the Heart Surface Potential (HSP) signal at the point where the EGM was recorded. There is a considerable risk of complications associated with such cardiac interventions. Medical professionals need to be experts and monitor all activity of movement of the catheter. EGM provides useful information about heart rhythm, power and functioning. Such investigations can identify many heart disorders, namely myocardial infarctions, heart failure, arrhythmia, and coronary artery disease. EGMs can also be used to direct cardiac medical treatments, like ablation therapy for arrhythmias and angioplasty for heart disease [3]. They may result in fatalities, heart attacks, strokes, and loss of human life [4]. Little children have small cardiac chambers and narrow veins, making this EGM treatment more risky when done on them [5]. Cardiac electrical instincts propagate throughout the body via particular tissues called cardiac muscle fibers [6]. When these (signals) impulses reach the skin's tissues, they can be measured using sensors called electrodes. This serves as the basis for electrocardiography (ECG), a non-invasive method of capturing and monitoring the electrical activity of the heart. The electrical activity of the heart's distinct chambers is shown in an ECG recording as a sequence of waveforms. The medical professionals place twelve sensors or electrodes at twelve different places of the body to capture the electrical signal of heart activity with 12 different angles. This procedure gives a clearer picture of the electrical activity of the heart as compared to a single lead ECG. Patients undergoing cardiac resynchronization therapy (CRT), a medical procedure that uses a pacer to synchronize the ventricles of heart, are frequently chosen based on twelve-leads electrodes ECG data. CRT is mostly helpful for patients with heart failure. Heart failure is a state in which the heart muscles are weak and unable to pump blood as efficiently as they can to supply oxygen and other compounds. However, around one-third of patients do not

have a favorable response to CRT treatment. This is most likely because ECG data is not precisely located and measures electrical inhomogeneities in the heart or due to a lack of expertise in ECG data understanding. Electrical inhomogeneities are zones of the heart where the electrical activities of muscles are abnormal.

Deep Learning (DL), a section of Machine Learning (ML), uses Artificial Neural Networks (ANNs) to learn from categorized data. The complex patterns in the data are recognized by ANNs. They are stimulated by the structure and operations of the human brain cells. Nowadays in a diversity of applications, such as Machine Translation (MT), Natural Language Processing (NLP), and Image Recognition (IR), DL gives the cutting edge results. These neural networks are used in industries like robotics, healthcare, finance, agriculture, and many more. In the 1940s when scientists started working on ANNs, they developed the DL concept and started exploring it. Still, it wasn't explored and used in real-world applications and problem-solving, until the early 2000s when DL received a lot of consideration, largely because of growths in learning algorithms and computational power. In 2006, Geoffrey Hinton published a research article [7] that presented how to train deep neural networks using a technique called backpropagation. This break-point opened the door for the conception of modern DL algorithms. In 2012, the first-ever ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [8] was won by a DL model. This event marked a crisis moment in the expansion of DL, and it proved the potential of DL to solve real-world problems. There is a wide diversity of DL methods that are used today. Some of the most common techniques of DL are discussed below.

An ML model type called an ANN is modeled after the composition and operations of the human brain. An elementary mathematical operation is carried out by each node that makes up an ANN. ANNs are used in DL, a kind of ML, to extract knowledge from data. DL models don't usually need to be explicitly coded to learn complex patterns from data. In the study [8], the authors tell a comprehensive review of relevant literature on Fuzzy Logic (FL) and ANNs in heart disease diagnosis revealing the potential of an accurate algorithm to save millions of lives worldwide. Moreover, it could enable remote care for critically ill patients, leading to cost and time savings, particularly in developing countries where healthcare facilities and infrastructure are unevenly distributed. Real-time monitoring of heart-related parameters could significantly enhance the quality of life and reduce the risk of heart disease.

Convolutional neural networks (CNNs) are a special kind of neural network that excels at tasks like photo identification that call for spatial (three-dimensional) input. Each convolutional layer's output is handled by the activation functions Rectified Linear Unit (ReLU), Sigmoid, Tanh, Maxout, Exponential Linear Unit (ELU), and Softmax. Batch normalization often improves CNN stability, and both can accelerate convergence speed [9]. Following convolutional layers with batch normalization and a ReLU activation function are frequently local pooling layers. These layers downsample the feature map, which lowers computing expenses [10]. Skin or other organic tissue injuries known as burns can result from exposure to various external factors, such as radiation, electricity, thermal energy, extreme cold, and chemical compounds. The extent and depth of tissue damage establishes the burn's severity. The degree of the injury must be taken into consideration while selecting a burn therapy [11]. Common treatment strategies that can improve outcomes for patients with

severe burns by reducing death rates and minimizing hospital stays include skin grafts, skin replacements, and early surgical removal of burned tissue (burn wound excision). On the other hand, delayed or insufficient care can result in adverse outcomes such poor wound healing, infections, discomfort, severe scarring, organ failure, and even death [12]. Burn injuries are usually divided into three categories by medical professionals: full-thickness (third-degree), superficial-partial (second-degree), deep-partial (third-degree), and superficial (first-degree). Every group is distinguished by distinct healing schedules and attributes [13]. Accurately determining the depth and severity of burn wounds at an early stage is extremely difficult because of their dynamic nature and propensity to get worse with time. The Deep Convolutional Neural Network (DCNN) architecture mixes transfer learning with fine-tuning to extract features from the images. It accomplishes this by stacking multiple convolutional layers over three different kinds of pre-trained models and adjusting their hyperparameters. Next, based on the intensity of the burns, a fully connected feedforward neural network is utilized to classify the images into first, second, and third degree burn categories [14]. Natural Language Processing (NLP) and other applications using sequential data are a good fit for recurrent neural networks (RNNs). The RNNs can use time-series data to extract sequential representations through a network of recurrent layers. Each recurrent layer is in charge of processing the relevant time step of the data. The hidden states of each layer are carried over the calculated values into the subsequent recurrent layer. Lastly, the last recurrent layer predicts the target class [15].

In 1997, Sepp Hochreiter and Jürgen Schmidhuber introduced the concept of Long Short-Term Memory (LSTM) in their paper titled “Long Short-Term Memory”. It was introduced to address the limitation of RNN. The “vanishing and exploding the gradients” problem during training, limits the RNN to capture the long term dependencies in the sequence [16]. Text, audio, and time series data are different types of sequences of data that LSTMs outshine at processing and foretelling. Scientists use NLP [17] in various complex tasks like speech recognition, and machine translation. A series of recurrently connected LSTM cells make up an LSTM network. Every LSTM cell consists of four major parts: The information from the previous cell that should be overlooked is selected by the forget-gate. The input-gate is in charge for selecting which new data or information to add to the cell state. Cell state: This is the memory of the cell, where it keeps records of the valuable information it has accumulated throughout time. The output gate selects the cell state information that will be the outcome of the cell state. The network of this new concept can discover long-term dependencies in data. The LSTM architecture authorizations information to be collapsable both forward and backward across the network. LSTM networks give cutting-edge results on a variety of tasks, including NLP [18], speech recognition [19], and machine translation.

Transformer neural networks have produced state-of-the-art outcomes in various applications, such as NLP, image recognition, and machine translation. In study [20], The authors apply a transformer network model for electronic health records (EHRs) that is intended to be scalable, interpretable, and tailored for a broad range of different diseases and EHR modalities. A pre-trained model named “Bidirectional Encoder Representations from Transformers (BEHRT)” can be adjusted for particular downstream tasks. To explore this feature, the

scientists trained and evaluated the model to predict the next most likely diseases in the future visits of patients on Clinical Practice Research Datalink (CPRD). BEHRT is adaptable enough to include more EHR data modes because of its sectional architecture.

For analysis of medical images, Deep Learning (DL) is being used to develop new algorithms. Through these algorithms, radiologists can more speedily and consistently diagnose situations and irregularities in medical pictures. For instance, DL models have been used to develop a system that can detect lung cancer in medical images with greater accuracy as compared to human radiologists [21]. DL is being used to speed up the drug discovery process.

Millions of potential drug candidates are screened by DL models for their safety and efficiency. These models help researchers to identify new treatments and drugs less response time. These models are used to develop new formulas for drugs against antibiotic-resistant bacteria [22]. Models create individualized treatment regimens for patients. The DL models can study inherited information, medical history, and other variables to predict the response of patients to various therapies. This helps models to design a personalized treatment plan based on the unique needs and behavior of the patient. Pancreatic cancer patient's individualized treatment regimens have been formed using DL models [23]. DL can be used to develop remote patient monitoring systems. The systems based on these models can be used to monitor vital signs, daily activities, and other health data of patients remotely. These large monitoring systems can help to identify early warning signs of health problems [24].

1.2 eXplainable Artificial Intelligence (XAI)

AI algorithms are capable of finding solutions to complex problems in a variety of disciplines. AI models sometimes lack interpretability and transparency, which makes it interesting to understand how they come to these decisions or target classes. This deficiency of explainability increases questions about the trustworthiness, accountability, and justice of the AI systems. The increasing adoption of AI models needs for eXplainable or interpretable Artificial Intelligence (XAI) [25]. The goal of XAI is to explain the inner workings and decision-making process of these models to their users. XAI can grasp the reasoning behind the predictions and how the model comes to the predicted class or classification. This technique is central for building confidence in AI systems decisions, ensuring accountability, and preventing potential biases or unfair outcomes. The idea of XAI can be dated to the initial phases of AI research [26]. Researchers in the 1980s, began discovering methods for explaining the outcome of expert systems, arranging the basis for XAI. These early efforts focused on rule-based structures, which were simple and easier to realize as compared to modern AI models. The development of ML algorithms and the increasing complexity of AI systems led to a change in interest in XAI [27]. Researchers started discovering methods for explaining the decision-making of ML algorithms like Decision Trees (DT), rule lists, and sensitivity analysis. These developed methods give some level of transparency but were frequently limited in their capability to explain complicated models. The field of XAI has progressive meaningfully in recent years, determined by the increasing density of AI models and the increasing demand for explainable solutions. Researchers have established a wide range of XAI

procedures, each with its strengths and boundaries. Some of the XAI tactics that have been established are shown in Figure 1.1. These are also discussed in some detail below.

1.2.1 Anchor Explanation

The type of XAI technique that provides local clarifications for discrete predictions is called Anchor explanations [28]. These explanations aim to categorize a set of features or data values that are most accountable for the predicted value of the model, acting as "anchors" that clarify the conclusion.

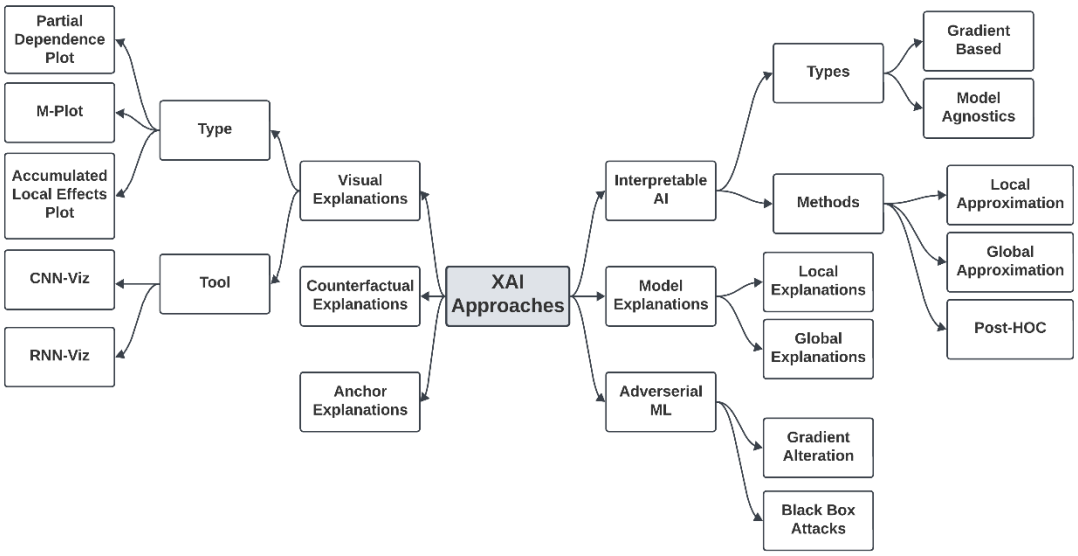

Figure 1.1: eXplainable Artificial Intelligence Approaches

Anchor explanations are beneficial for explaining AI models, like Deep Neural Networks (DNN), where it can be hard to recognize the influence of distinct features.

1.2.2 Visual Explanation

Visual explanations are a powerful tool that is used for communicating XAI insights to users [29]. They plot graphic components or items like charts, graphs, and images to display the complex associations between features and calculations. These tools generate more intuitive and easier to understand as compared to text-based explanations. Both explanations, local and global, can be better understood with the use of visual assistance. These visual assistances provide light on precise prediction values as well as the universal behavior of the model.

1.2.3 Counterfactual Explanation

Counterfactual explanations [30] are a type of XAI technique that provides explanations by generating alternative input scenarios that would lead to a diverse forecast from the model. This can assist users in understanding how the model's forecast could have been altered by making various choices in the input data. Counterfactual explanations [31] are particularly useful for identifying potential biases in AI models, as they can highlight how the model's predictions are influenced by certain features or data points.

1.2.4 Interpretable AI

Interpretable Artificial Intelligence (IAI) is the study of developing AI models with fundamental interpretability [32]. That means, their decision-making procedures can be understood without the requirement for exterior explanation techniques. This is an interesting goal, as it requires the progress of new Artificial Intelligence algorithms [33] which are designed to be transparent and explainable from the outset.

1.2.5 Adversarial ML

Adversarial ML is also a field of XAI in which scientists develop methods and techniques for attacking and manipulating AI models. The skillfully constructed inputs that deceive AI models into generating inaccurate predictions is an example of adversarial ML. It is a powerful tool that is also used for testing the robustness and security of AI-based complex systems [34]. It can be utilized to develop new XAI techniques by identifying how AI models can be deceived and manipulated.

XAI can offer numerous paybacks, including *Quality of understanding of AI models:* Researchers may make improved decisions and grasp how AI models work with the help of visualizations. *Increased trust in AI systems:* By providing transparency, graphic explanations can help to build confidence in AI systems. *Identification of potential biases:* Graphical representation helps to find possible biases in AI models by prominence patterns in the dataset that may not be seeming from text-based clarifications. *More informed decision-making:* Graphic explanations can help users make more knowledgeable conclusions by providing intuitions into the features that impact the predicted value/s.

1.3 Problem Formulation and Problem Statement

In the modern field of Artificial Intelligence, many models are being established to assist in the medical industry. Many machine learning and deep learning models are being created to support physicians and other health professionals in disease prevention, and estimate. These models also help in learning how to decrease the probabilities by advising diets and other supplements. Many scholars have already conducted research on XAI and heart disease prediction. Some issues which this study focused are mention below.

1.3.1 Lack of Sensitivity to Minority Class

The main problem is that the model is not very good at identifying instances of the minority class. The poor recall score, which reflects the imbalance, suggests that the model has difficulty capturing positive cases and may result in considerable false negatives.

1.3.2 Designing an Attention-Based MLP Model

This study intends to develop an attention-based MLP model which gives better accuracy, better recall, precision, and other evaluation measures.

1.3.3 XAI Model Analysis

In this study, XAI Post-Hoc interpretable methods are also used for visual understanding, Model clarifications, feature relevance explanation, and picture with simplification. This exploration will assist doctors and other

medical personnel in understanding the aspects that should be within normal ranges to improve the quality of health and life.

1.4 Research Objective

This study addresses the challenges of working with class-imbalanced data. It highlights how irregular class distribution can affect model prediction behavior when feeding real-world feature values. This approach covers preparing the data with data augmentation, refining the attributes and flags of models, and using explainable AI tools like SHAP, to analyze to improve understanding. The area of this study is to make models more reliable and effective when dealing with real-world data.

1.4.1 Addressing Class Imbalance

It is important to use different approaches to overcome the class-imbalance. Techniques like oversampling the class that have less samples to meet the equal samples of other classes. The study uses the Python "imbLearn" library module named "SMOTE" to make the dataset more balanced. The proposed model works great when it is trained on a balanced dataset.

1.4.2 Fine-Tuning Model Architecture

The capacity of the model to classify patterns in both balanced and unbalanced dataset can be enhanced by iteratively fine-tuning the model architecture, and exploring various possibilities for the number of hidden layers, number of nodes in different layers, number of epochs, and batch size.

1.4.3 Utilizing SHAP Analysis

Using SHAP algorithms (XAI tool) benefits recognizing how each feature impacts the prediction of the model in detail. The visualization from SHAP analysis can assistance expand feature selection and polish the model. The SHAP algorithm gives the feature ranking and also gives the individual feature values impact on the prediction.

This research study will increase trust in artificial intelligence systems and establish the transparency and understandability of artificial intelligence systems in the health department.

1.5 Research Contribution

This research study points out the challenges in imbalanced datasets. This aims to explore these tasks in-depth, propose new resolutions, and add valuable understandings to the field. The research discovers new areas and builds a strong groundwork for further studies. The results give important visions, corroborative existing knowledge and reveal new conclusions. The novelty in the methods used to discover the patterns in imbalanced data creates a meaningful influence. Hypothetical perceptions force current frameworks further, providing deeper thoughts about how these models handle data and predict the target class. This research study fills important knowledge gaps and challenging and growing current debates. It meaningfully affects future research, improving both theory and real-world applications.

1.6 Thesis Outline

The categories that make up the thesis are as follows. Chapter 2 presents the best prior research papers on the use of AI for heart disease diagnosis or prognosis. The dataset's exploration, a suggested DL model, the SHAP framework, and performance indicators are all explained in Chapter 3. In Chapter 4, the outputs of the dataset are given and discussed. Chapter 5 concludes with a summary of the suggested research project.

Chapter 2: Literature Review

Recent research is summarized in this chapter. Cardiovascular disease, which stands as the predominant global cause of mortality, presents a formidable obstacle to healthcare infrastructures universally. The emergence of AI, particularly through the avenues of ML and DL, has brought about a paradigm shift in the realms of heart disease diagnosis, prognosis, and management. AI algorithms exhibit the capacity to scrutinize extensive datasets encompassing patient medical histories, clinical metrics, and imaging outputs. Through discerning patterns within this data, these algorithms can render predictions with exceptional precision. Harnessing the predictive capabilities of AI enables healthcare providers to prospectively identify individuals at heightened risk of developing heart disease, facilitating timely interventions and the formulation of personalized therapeutic strategies.

2.1 Machine Learning Models

H. Ahmed et al [35] provide a novel real-time approach to the proactive prediction of heart illness based on ongoing medical data streams that reflect a patient's current state of health. Finding the finest ML system that can accurately forecast cardiac problems is the main objective. To improve predictability, two independent feature selection algorithms are used to extract critical features from the dataset: univariate feature selection and Relief. A comparison of four prominent ML methods is performed: DT, SVM, RF, and LR. The evaluation is carried out utilizing both selected features and the entire feature set the system's accuracy is refined through the application of hyperparameter tuning and cross-validation techniques. Empirical findings highlight the RF Classifier as the most effective model, achieving the highest accuracy rate of 94.9%.

In J. Rashid et al [36], the authors employed classification techniques such as CatBoost (CB), decision tree (DT), XGBoost (XGB), bagging (BG), AdaBoost (ADA), and the proposed new algorithm. The authors achieved 93% accuracy, 89% sensitivity, and 96% specificity using the proposed methodology.

Mohan et al [37] introduce an innovative approach, embedded in the IoT landscape, to heighten the accuracy of predictions. The proposed model, incorporating diverse feature combinations and established classification techniques, achieves an impressive accuracy value of 88.7% through the Hybrid Random Forest with a linear model (HRFLM). Emphasizing the pivotal role of processing raw healthcare data for heart information, the study underscores its potential for long-term life-saving and early detection of cardiac anomalies. While acknowledging the intricacies of heart disease prediction within the medical realm, the study underscores the impact of early detection and preventive measures in curbing mortality rates. Advocating for an extension into real-world datasets beyond theoretical frameworks and simulations, the study validates the efficiency of the HRFLM approach, fostering future exploration of diverse ML techniques and innovative feature-selection methods to enrich comprehension of significant features and refine predictive performance.

Soni et al [38] conduct a comprehensive survey of contemporary knowledge discovery techniques, specifically applied to subject disease within medical research. The study explores experimental comparisons of predictive data mining techniques in recognition of the lack of useful analysis tools for revealing hidden links in healthcare data. Notably, DT demonstrates greater performance than k-nearest neighbor (KNN), ANNs, and cluster-based classification. The research highlights the enhanced accuracy achieved by DT and NB when

complemented with genetic algorithms, strategically optimizing attribute subsets crucial for heart disease prediction. The focus extends to diverse algorithms and target attribute combinations, aiming for intelligent and effective heart attack prediction with a list of 15 significant attributes. To broaden the scope of forecasting methods, the study promotes the possible integrations of other approaches such as ANNs, Time Series, Clustering, Association Rules, and soft computing.

Ali et al [39] recognizes the paramount importance of accurate disease prediction, showcasing the remarkable performance of ML approaches. Notably, the KNN, DT, and RF algorithms exhibit stellar accuracy, reaching 100% on a heart disease dataset. Feature importance scores, meticulously examined except for MLP and KNN, offer valuable insights into the relevance of different features. The conclusion affirms that these ML techniques, renowned for their wide acceptance and ease of implementation, present promising outcomes, representing an initial stride in incorporating ML approaches for advanced patient care.

Shah et al [40] use supervised learning algorithms including NB, DT, KNN, and RF and concentrate on characteristics linked to heart disease. Using a dataset of three hundred three (303) instances and seventy-six (76) attributes from the UCI repository's Cleveland database, the research carefully evaluates fourteen (14) attributes in order to validate algorithmic performance. The primary objective is to envision the probability of heart disease development, with KNN demonstrating the highest accuracy 90.789%. The main objective is to specify effective data mining methods with an emphasis on accuracy using a reduced set of variables for accurate heart disease prediction. The paper recommends more research to overcome constraints and improve predictive accuracy for early cardiac disease identification, using new data mining techniques such as SVMs and time series analysis.

Bhatt et al [41] aims to develop a predictive model for subject diseases, proposing a k-modes clustering method with Huang starting to enhance prediction accuracy. Various models, including RF, DT, MLP, and XGBoost (XGB), are employed and optimized using GridSearchCV. Utilizing a real-world dataset of 70,000 cases from Kaggle, the models show accuracies ranging from 86.37% to 87.28%, with corresponding Area Under the curve (AUC) values of 0.94 to 0.95. The MLP with cross-validation outperforms other algorithms, attaining the maximum accuracy of 87.28%. K-modes clustering is applied to a heart disease patient dataset, preprocessing it by age conversion and blood pressure binning. Gender-based dataset splitting considers unique disease progression. The elbow curve method determines optimal clusters, revealing the MLP classifier's maximum accuracy of 87.23%. These outcomes underscore k-modes clustering's potential for precise heart disease prediction, suggesting its utility in targeted diagnostic and treatment strategies. Despite promising results, limitations include dataset specificity, an absence of broader risk factors consideration, lack of test dataset evaluation, and unexplored cluster interpretability, warranting further research to address these aspects and elucidate k-modes clustering's potential in heart disease prediction. Determining optimal features for ML models is a complex task, particularly in predicting CVD accurately.

In the research [42] ,the Pandita create a time and cost-efficient system capable of accurately determining the presence of subject disease. Among the ML algorithms assessed, KNN emerges as the most effective in model,

achieving an accuracy of 89.06%, while LR yields the least accurate prediction at 84.38%. The deployment of such advanced algorithms holds significant promise in the proactive management of heart disease, aligning with the imperative to minimize its prevalence and impact on a global scale.

Lakshmanarao et al [43] delves into the intersection of medicinal science and information mining to explore metabolic disorders. Employing ML, a technique enabling systems to learn from historical data without explicit programming, proves pivotal in heart disease detection, demonstrating its impact on enhancing accuracy and recall rates. The research utilizes ML methods for heart disease detection, addressing class distribution imbalances in raw datasets through three distinct sampling techniques. Results showcase significant accuracy improvements, with SVM achieving the highest accuracy of 99.0% for random oversampling, while SMOTE sees RF and Extra-Tree Classifier attain the best accuracy at 91.3%. Adaptive synthetic sampling yields commendable accuracy, with RF and Extra-Tree Classifier reaching 90.3%. This study underscores the efficacy of ML approaches, particularly when addressing class imbalances, in advancing heart disease detection accuracy.

In [44] A comprehensive literature review noted that the mainstream of studies focused on the Cleveland dataset, which is characterized by a limited 303 instances and 14 features, restricting its representativeness of specific geographic areas. This prevalent use of a single dataset across studies poses challenges in generalizing classification accuracies for heart disease prediction. To address this limitation, future research endeavors aim to explore multiple datasets of heart disease from diverse geographic algorithms with increased dimensions, aiming for more generalized and efficient ML models. The ongoing research is fundamentally driven by the goal of achieving enhanced classification and early prediction of heart diseases, ultimately mitigating the rising rates of morbidity and mortality associated with CVDs.

The research study led by Alotaibi [45] contains a comprehensive examination of subject disease classification and prediction by applying ML techniques. The study applies algorithms like NB, DT, RF, SVM, and LR in the RapidMiner framework. The study utilizes the broadly accepted Cleveland heart disease dataset from the UCI repository. The dataset comprises 303 instances and 14 attributes. The learning and evaluation of the model employ the 10-fold cross-validation method. Results show that the DT algorithm shows the peak accuracy in subject disease prediction, followed by SVM at 93.19% and 92.30%, respectively. The research study presents a model joining five algorithms in the RapidMiner tool, representing higher accuracy than the Matlab and the Weka software. Despite the limits of a small dataset, the research study shows important enhancements over previous research studies, highlighting the potential utility for timely diagnoses in the medical field.

Bora et al [46] report the challenge through the scheme of heart disease prediction applying various ML algorithms, including LR, NB, SVM, KNN, RF, extreme gradient boost, etc. The research study uses two diverse datasets, one from the UCI ML repository with 303 records and 14 attributes, and another dataset from Kaggle with 1,190 patient records and 11 features. The amalgamated dataset, a fusion of five popular datasets, simplifies a complete estimation of ML techniques. The results show a peak accuracy of 92% using the SVM

algorithm for the UCI dataset, 94.12% with RF for the Kaggle dataset, and an overall peak accuracy of 93.31% using RF on the joint dataset. These results highlight the efficiency of ML algorithms in enhancing heart disease prediction accuracy, flagging the way for more active preventative measures in the medical domain.

Ayatollahi et al [47] answers the utility of data mining algorithms, particularly the Support Vector Machine (SVM) model, in predicting the CAD. The SVM model revealed higher evaluation metrics, including Lower Mean Absolute Percentage Error (LMAE), a delicate Hosmer-Lemeshow test result (16.71), and higher sensitivity at 92.23%. The feature triggering the CAD showed a better fit in the SVM model than the ANN model. The ROC curve of the SVM algorithm also confirmed its higher accuracy. The research study recommends more research to compare different ML algorithms and find the best one for predicting diseases. The SVM model stands out for its accuracy, and performance to predict CAD.

Rindhe et al [48] thoroughly explore existing techniques, aiming to identify efficient and accurate systems. ML emerges as a transformative force, significantly enhancing the accuracy of cardiovascular risk prediction. Subsequently, three models were trained and tested, achieving maximum scores as follows: SVM: 84.0%, Neural Network: 83.5%, and RF: 80.0%. The results contribute to predicting treatment strategies for patients. This improvement enables early disease identification, facilitating timely preventive treatment for patients.

In a cross-comparative study, Shorewala [49] uses a risk factor approach and learning strategies including KNNs, Binary Logistic (BL), and NB. By adding randomness to the data, K-Fold validation evaluates the consistency of the model's output. Furthermore, hybrid (mixture) models are examined, which combine cross-comparisons with basic classifiers and ensemble methods like as bagging, boosting, and stacking. The Cardiovascular Disease Dataset, which consists of 70,000 records of medical examinations for subject disease, is used to test the algorithms. The accuracy of bagged models is on average 74.8% higher than the traditional equivalents. Boosted models have the highest AUC score of 73.0 and an average accuracy of 73.4%. With an accuracy of 75.1%, the stacked model, which combines KNN, RF classifier, and SVM, proves to be the most efficient.

An intelligent diagnostic system [50] for heart diseases has been developed to address the potential misdiagnosis issues encountered by medical professionals. Utilizing the Statlog Heart Disease dataset that was acquired the UCI ML repository, the experiment focuses on attributes associated with patients diagnosed for subject disease, aiming to confirm the presence or absence of the condition. The dataset is strategically divided into training, validation, and testing subsets for effective model training. The intelligent system is implemented using feedforward multilayer perceptron and SVM models. Comparative analysis of the recognition rates reveals that the SVM outperforms, yielding a recognition rate of 87.5%, while the feed forward multilayer perceptron achieves 85%. This experiment concludes that the SVM model stands out as the optimal choice for heart disease diagnosis in the medical field.

P. Ghadge et al. [51], the authors suggested a system for mining unseen knowledge (correlations and patterns) related to disease from already developed heart disease database system. Authors use of software like Hadoop, a Java framework for distributed processing and storing of big datasets. A. Rajkumar and M. G. S. Reena [52],

Using data mining, the scientists produced a system. This study's approach shows the assistance of health practitioners in making timely correct decisions with the help of input patient data. In the training process proposed system using the 10-fold method. This approach discovered an accuracy value of 87.0 % in the training and the testing phase achieves 86.0 %. Through this approach the model gives better results and assists experts and even persons associated with the health department to make for a greater adjustment and give the patient purpose to fight back with the disease.

A. Hazra et al. [53], the authors worked on the diagnosis of cardiac disease using supervised ML classification. An AI tool named *Tanagra* is utilized to classify the dataset, and cross-validation with 10-fold is used to gauge the dataset before comparing the findings. This AI tool is a free data mining Windows OS based application for educational and research use. It recommends many data mining techniques from the fields of clarifying data analysis, the statistical learning, ML, and database management. The dataset splitted into two subsets: training set 80% and testing set 20%. They developed and designed an evolving neural network for detecting subject disease. This study offers a new system for detecting cardiac problems that employ the most famous feed-forward neural structure and Genetic Algorithm (GA). The suggested approach intends to make cardiac disease diagnosis easier, more cost-effective, and more reliable. The dataset collected from the *University of California*, *Irvine database*. The weights of the nodes in the ANN with 13 input nodes, then two hidden layers, and 1 output node are adjusted using Gradient Descent (GD) and then GA. In the study, authors compared the different methods, and it is concluded that the GA can choose the ideal weights efficiently. In a GA, picking one individual from a population of people the Tournament selection function is used. According to this study, the authors choose more members of offspring population. It is a sign of the development of offspring, which leads to increased diversity and survey of the population.

The use of unstructured EHRs as possible repository for automated evaluations of patients' 10-year risk of Coronary Artery Disease (CAD) is explored by Jonnagaddala et al [54]. Using appropriate imputation techniques, the study tackles the problem of missing data in these records. To compute 10-year CAD risk scores from shapeless EHRs, a text mining with rule-based method is presented, exhibiting the ability to extract a significant amount of documented risk factor data. The system calculates Framingham Risk Scores (FRS) for 164 eligible patients, acknowledging that not all patients possess the requisite risk factor data. Despite this, the scores generated align consistently with manually calculated scores. Results reveal a prevalent FRS between 10% and 20%, attributed to the cohort's diabetic nature. Innovative data exploitation from a corpus originally created for various purposes is highlighted in the study, and methodologies relevant for risk stratification and cohort discovery in studies requiring FRS calculation from unstructured EHRs are highlighted. The flexibility of the text mining technique to extract non-framingham risk factor data is highlighted for building CAD prediction models. The study outlines future directions, including corpus annotation for performance evaluation and a planned comparison of scores calculated using proposed methods against clinician manual determination.

The conventional usage of classification trees for patient categorization based on disease presence encounters accuracy constraints, prompting exploration into alternative methods within the data-mining and ML spheres. effectively compares the performance of contemporary, adaptable tree-based techniques, such as bagging, boosting, RFs, and SVMs, with the effectiveness of classical classification trees in the specific classification of two Heart Failure (HF) subtypes: HF with decreased ejection fraction and HF with Preserved Ejection Fraction (HFPEF). Furthermore, these approaches' predictive power for calculating the likelihood of HFPEF is contrasted with traditional logistic regression. Results highlight the substantial enhancements provided by contemporary tree-based methods in predicting and classifying HF subtypes when in contrast to regression trees and traditional classification. Notably, logistic regression outperforms the proposed data-mining literature methods in forecasting the probability of HFPEF. In a sample from Ontario, Canada, tree-based methods excel in HF subtype prediction while demonstrating comparable performance to logistic regression in forecasting the presence of HFPEF. In the two-stage feature subset retrieval technique, Hasan and Bao [55] take into account three well-known feature selection techniques (filter, wrapper, and embedding) and extract a subset based on a common “True” condition that is driven by a Boolean process. Using ANN as a benchmark, RF, SVM, KNNs, NB, and XGB models are used to evaluate comparative accuracy. The results show that the innovative component of common “True” condition-based feature selection in medical informatics is provided by the XGB Classifier linked with wrapper approaches, which yields exact predictions for cardiovascular illness. A multi-stage learning algorithm for feature selection by resampling is introduced in this paper. Feature selection plays a crucial role in streamlining the learning process for prediction models in cardiovascular disease. Experimenting with a dataset of 70,000 patient records, the top ten significant features include weight, BMI, ap_lo, age, height, ap_hi, gluc, cholesterol, active, and alco, with XGB and SVC as the top-performing classification models. Limitations include a low-dimensional attribute set and a single dataset, suggesting opportunities for future research to explore high-dimensional datasets, incorporate diverse cardiovascular datasets for comparison, explore alternative dimensionality reduction techniques, and apply the multi-stage learning algorithm in other domains beyond healthcare.

C. S. Dangare and S. S. Apte’s research [56] takes a significant stride by incorporating two crucial input attributes (obesity and smoking). This addition aims to propel the accuracy of research predictions to new heights. The authors employ three formidable data mining classification techniques (DTs, NB, and Neural Networks) to unravel the intricate patterns within dataset. Among these, Neural Networks emerge as the unsung hero, consistently delivering more accurate predictions compared to its counterparts, DTs and NB. By providing a platform for the integration of several data mining methods, such as Clustering, Sequence of Time, and Association Rules, the system they built lays a solid foundation for future growth. Moreover, the prospect of delving into text mining opens up new possibilities, allowing us to glean valuable insights from the wealth of unstructured data nestled within the healthcare industry's expansive database.

Using a variety of rule mining techniques, including Apriori, Predictive Apriori, and Tertius, Nahar et al [57] conducted a thorough investigation into the extraction of rules from data related to heart disease. Gender-

specific stratification of the data also reveals unique risk variables for men and women. One noteworthy discovery from the examination of healthy rules is that there is a correlation between being “female” and having a good heart condition. This suggests that women are more likely than men to not have the CAD. This research study aligns with other medical research showing that premenopausal women have a lower risk of having this disease. It is due to the protective effects of estrogen. The research study also explores how lack of iron in younger women bodies, caused by menstruation, might help delay the beginning of heart disease. Using rule mining, the research study reveals key insights, pointing out that factors like chest pain and exercise-induced angina signal heart disease in both men and women. Gender differences are detected, with resting-ECG being a key factor in forecasting the CAD for women, while for men, an abnormal resting-ECG is allied to higher risk. Both men and women share healthy signs like an upward slope and an oldpeak value of 0.56 or lower, indicating good health. Beyond these findings, the study expresses the importance of rule mining and AI in classifying illness factors and addressing the problems in medical research. It emphasize how computational approaches assistance understand subject disease factors across genders.

The study of Leila Baccour [58] begins with the growth of the Amended fused TOPSIS-VIKOR methods for classification (ATOVIC), which smartly participates the Multiple-criteria Decision Making (MCDM) approaches of VIKOR and TOPSIS into the classification method. Departing from orthodox MCDM norms, ATOVIC schemes into unexplored territory, approval of three sets: classes, objects, and attributes (features). The research shows that criteria become features, and alternatives change to objects linked to precise target classes. A thorough test on the CLEVELAND dataset highlights ATOVIC as a top performer in forecasting of subject disease. It shows asset in both binary and multi-class classification, showing its flexibility. ATOVIC also shines in predicting thyroid diseases, outdoing as compared to the other classifiers. It continues to rule across different datasets like chess, nursery, and titanic. The study suggestions at future plans to integrate type-1 fuzzy logic and its advanced methods, like intuitionistic and type-2 fuzzy logic to explore more. Moreover, the horizon reveals an alignment with contemporary trends, as ATOVIC gears up to navigate the realms of Big Data, echoing the trajectory of stalwart classification tools like SVM and ANN. The tale of ATOVIC unfolds as a compelling chapter in the ever-evolving saga of classification, where innovation meets adaptability on the frontier of scientific exploration.

Using a combination of descriptive and predictive techniques, Shamsollahi et al [59] analyzes 282 patient records with 58 parameters that were taken from a clinical dataset in order to predict CAD. The right number of clusters 3 was established based on clustering indices by using the k-means clustering method for descriptive purposes and different classification methods (CHAID, Quest, C5.0, C&RT DT, and ANN) for prediction. Then, DT techniques were used on every cluster, exposing unique features. With a 0.074 error, C&RT was shown to be the most efficient strategy overall for the full dataset. Notably, the optimal prediction method varied for each cluster, with C&RT performing best. The proposed procedure carries clinical implications, suggesting the development of user-friendly software in heart clinics for CAD diagnosis, utilizing the combined method's results. This research introduces a model integrating descriptive and

predictive data mining techniques to enhance CAD prediction in healthcare systems, showcasing the efficacy of C&RT as the optimal method for overall accuracy. See the summary in Table 2.1.

Table 2.1: Summary of Machine Learning Models

Ref	Method	Dataset	Accuracy %
[35]	Heart illness detection from patients' social media posts using DT, SVM, RF, LR	Hungarian	94.90
[36]	Brute Force Algorithm for feature extraction of heart disease along with NB, RF, SVM, and KNN	Statlog, Cleveland, and Hungarian	94.0
[37]	Hybrid ML techniques (DT, SVM, RF, NB, NN, KNN) for effective CVD prediction	Cleveland	88.7
[38]	Predictive Data Mining with DT, KNN, and ANN for Medical Diagnosis	StatLog	93.25
[39]	ML algorithms KNN, DT, RF	Framingham	93.2
[40]	ML Methods NB, DT, RF	Cleveland	90.78
[41]	ML techniques K-Mode Clustering, RF, MP, XGB	Cardiovascular dataset	87.2
[42]	ML Algorithms KNN, LR	Cleveland	89.06
[43]	ML Methods SVM, RF	StatLog	99.0
[44]	Data-mining and ML Methods (NB, DT, and ANN)	Cleveland, Statlog	86.6
[45]	ML Model (RF, DT, SVM, LR for Heart Failure Prediction	StatLog	93.19

2.2 Deep Learning Models

Improving the outcome of cardiac disease requires an early diagnosis and timely treatment. However, the requirement for huge datasets hinders the effectiveness of current automated diagnostic techniques. Al-Makhadmeh and Tolba [60] proposes an Internet of Things (IoT)-based medical gadget that gathers extensive cardiac data from patients both before and after the development of the condition. A Higher Order Boltzmann Deep Belief Neural Network (HOBDBNN) is employed to handle the data after it is sent to a medical facility. This DL technique leverages knowledge from prior analysis to extract pertinent features related to heart disease. The system's efficiency is improved through the use of complex data structures. Utilizing the f1-score, specificity, sensitivity, ROC curve, and loss function in experimental evaluations, the system maintains a low time complexity of 8.5 s while achieving an accuracy of 99.03%. This innovative approach significantly reduces the complexity of heart disease diagnosis and has the potential to lower heart disease mortality rates. Akella and Akella [61] focus on the application of ML techniques for CAD prediction in patients. The outcomes validate the accuracy of ML algorithms in CAD prediction. Publicly sharing the code aims to enhance ML algorithms' diagnostic utility for CAD. The consideration of employing the SMOTE methodology

to generate synthetic data resulted in improved accuracy, although concerns about the authenticity of these synthetic data points led to the decision not to implement SMOTE. Utilizing various ML algorithms, accuracy consistently exceeds 84.0%, with outstanding performance from the neural network model, achieving 93.03% accuracy and 93.80% recall. Multiple experiments with varying training and test set proportions affirm the neural network's consistent performance. Excluding accuracy from mean calculation due to its potential misrepresentation in biomedical datasets, other models exhibit high accuracy values: RF 87.64%, Generalized Linear Model 87.64%, SVM 86.52%, and KNN 84.27%. These findings contribute valuable insights into ML algorithms' efficacy for CAD prediction.

Das et al [62] provide an approach that uses SAS Base Software as its engine and is based on an ensemble neural network technique. SAS Base is a fourth-generation programming language for data access, data transformation, analysis, and reporting. This proposed technique, which is the foundation of the methodology, coordinates the combination of following possibilities or expected values from several classifier models to create models with increased effectiveness. With the Cleveland Heart Disease Database as a backdrop, the experimental journey achieves an impressive 89.01 classification accuracy. With sensitivity and specificity values of 80.95 and 95.91, respectively, the technique is highly effective in diagnosing cardiac disease. The ensemble model, a masterpiece woven from three independent neural network models, exhibits its mettle, with attempts to amplify performance falling short of improvement. SAS Enterprise Miner 5.2 emerges as a linchpin, seamlessly supporting all requisite tasks and providing a flexible canvas for collaborative endeavors. Its strong features cover performance evaluation tests, give operators a panoramic sight of the system. Medical progression is established by the development of Carotid Artery Stenting (CAS) as the main action for cerebrovascular stenosis. But this capable avenue is not without its challenges, mostly for older patients who face the threat of Major Adverse Cardiovascular Events (MACE).

In response, Cheng and Chiu [63] developed an Artificial Neural Network (ANN) model to predict the forecast of CAD, using data from 317 patients from the Taiwan National Health Insurance Research Database (TNHIRD) [64]. The proposed model in this research study was trained and tested with 13 clinical risk factors as input and MACE occurrence as the output. An MLP with 18 neurons in its hidden layer, achieved 89.4% sensitivity, 57.4% specificity, and 82.5% accuracy in evaluating process. For the entire dataset, it maintained good performance with 85.8% sensitivity, 60.8% specificity, and 80.76% accuracy. This model not only forecasts the target class but also assists in identifying high-risk CAD patients, helping communication between the neurologist doctors and cardiologist doctors for better treatment decisions.

The Back Propagation Neural Network (BPNN) appears as a mostly effective method for categorizing the hypertension gene sequences, representing a remarkable 90% precision rate for a smaller batch size of 80. After exploring the target gene sequences, Zaman and Toufiq [65] propose a modern method for organizing hypertension gene sequences using BPNN. They complete this challenge by using the frequencies of codons, or nucleotide triplets, as an individual metric. Using a range of sample sizes, the study systematically inspects how well the BPNN style approach achieves the goal during the training and testing stages. The findings show

that accuracy increases proportionately with sample size, which suggests that the BPNN classifier's classification error rate is declining. The definition of gene sequences is often achieved through Single Nucleotide Polymorphism (SNP), amino acid, protein, and mutation synthesis; however, this work effectively applies a novel strategy called codon frequency. Codon bias, or frequency, proves to be a robust parameter for the classification of hypertension gene sequences. Unlike traditional BPNN applications that focus on predicting hypertension solely from physical characteristics or risk factors, this study is a pioneer in the creation of a BPNN system for codon classification of gene sequences-based hypertension prediction. Any sequence can be categorized to find its relationship to the disease after the training phase. Furthermore, prediction rates can be computed because gene sequences are accountable for the illness. This groundbreaking approach shifts hypertension diagnosis from reliance solely on phenotype to the integration of genotype or gene sequences, paving the way for comprehensive and early prediction of hypertension along with pre-diagnosis.

The heart disease prediction system presented in this study [66] by Subhadra and Vikas utilize a MLP, employing the backpropagation algorithm for effective training and iterative parameter comparison. The iterative nature of the backpropagation algorithm ensures the attainment of minimal error rates, resulting in maximized accuracy rates, as evident from the presented results. The proposed methodology expressions better efficiency in forecasting the CAD using 14 attributes related to other research methodologies. Diagnosing heart disease needs careful examination of patient medical test results and health history. Advances in machine learning algorithms, particularly in implementing the smart automated systems, provide useful tools for medical doctors to predict and come to the decisions about target diseases. The proposed system could expand timely medical care and assistance save lives. Meshref et al told a comprehensive analysis of the Cleveland heart dataset in research study [67], using ML classifiers to improve diagnostic models. For example, the MLP model reached 84.25% accuracy but with an 8-feature set, rising concerns about its correctness for subject disease recognition. The investigation is complemented by an interpretation analysis introducing the Feature Ranking Cost (FRC) index, a useful measure that makes it easier to distinguish across models according to the significance of their feature sets. The ultimate choice, the RF model, with 79.92% accuracy, finds a better way to balance accuracy and transparency than the MLP model, making it a more genuine choice. This research addresses the need for interpreting ML models, a vital aspect often overlooked in favor of high accuracy. ANN exhibiting the highest accuracy at 84.25%.

Romdhane et al [68] introduce a novel DL approach utilizing a CNN model. CNN models may automatically perform feature extraction while the classification process is ongoing, removing the requirement for a separate feature extraction step using human methods. To set it apart from other approaches, the technique also uses a unique heartbeat segmentation algorithm. Every ECG heartbeat is started at an R-peak by this segmentation technique, which ends after 1.2 times the median RR time interval in a frame of 10 seconds. The simplicity and effectiveness of this approach lie in its absence of signal morphology or spectrum assumptions, without resorting to filtering or processing. Even with earlier attempts to develop better algorithms for classifying ECG

heartbeats, study results were not ideal, especially when the datasets were unbalanced. In response, the authors introduce an optimization phase that employs a novel loss function called focused loss in conjunction with the deep CNN model. By emphasizing minority heartbeat classes, this function focuses on them. The model showed improved performance after being trained and evaluated on the MIT-BIH and INCART datasets for the aim of identifying the Association for Advancement of Medical Instrumentation (AAMI) standard's five arrhythmia classifications (N, S, V, Q, and F). Overall findings showed 98.41 recall, 98.38 F1-score, 98.37 precision, and 98.41% overall accuracy. Moreover, the technique performed better than current cutting-edge techniques.

Dutta et al [69] introduce a neural network with convolutional layers designed for efficient classification of highly class-imbalanced clinical data, particularly derived from the National Health and Nutritional Examination Survey (NHANES) to predict coronary heart disease (CHD) occurrences. Different existing AI classifiers susceptible to class imbalance, two-layer CNN exhibits resilience, achieving a harmonious balance in class-specific performance. A two-step strategy is employed: firstly, authors utilize LASSO-based feature weight assessment and majority-voting to identify crucial features, followed by homogenization through a fully connected layer. Authors propose an epoch-wise training routine, akin to simulated annealing, enhancing classification accuracy. Despite NHANES dataset imbalance, proposed CNN attains 77% accuracy for CHD presence and 81.8% for absence on testing data, indicating generalizability to similar healthcare studies. Compared to SVM and RF, proposed model demonstrates superior negative case prediction accuracy, offering potential for enhanced medical diagnostics and reduced costs in healthcare systems. With a balanced accuracy of 79.5%, the CNN outperforms individual SVM or RF classifiers, exhibiting high specificity, test accuracy, recall, and AUC values.

Du et al [70] use big data and ML to create an accurate model for CHD prediction that targets a significant number of hypertension patients in Shenzhen, China. The authors used electronic health records from 42,676 individuals, 20,156 of whom had CHD at onset, throughout a period of 1~3 years before to beginning or over a followup period of more than 3 years without any disease. To construct appropriate prediction models, the selected dataset was divided into distinct training and test subsets. The training set was subjected to a variety of ML techniques. For the independent test dataset, the XGB ensemble technique showed excellent accuracy in predicting the onset of 3-years CHD, with an Area Under the Receiver Operating Characteristic (AUROC) curve value of 0.943. Comparative studies exhibited that ML techniques performed better than conventional risk scales and that nonlinear models (RF AUC 0.938, KNN AUC 0.908) were superior to linear models (LR AUC 0.865). Further studies showed that, in comparison to utilizing solely static characteristics, including time-dependent data from numerous records, that is, statistical and changing-trend variables, improved model performance. Subpopulation analysis highlighted the effect of feature design on model accuracy, showing highly nonlinear characteristics regarding risk scores for both traditional and EHR components. Furthermore, the accumulation of EHR data over several time periods offered useful characteristics for improved risk prediction, highlighting the importance of gathering big data from EHRs to improve illness forecasts.

Neural networks have gained prominence for enhancing accuracy. However, dissatisfaction arises among medical experts due to the inherent "black-box" nature of deep neural networks. In response, Kim and Kang [71] introduce an NN-based CHD risk prediction model employing Feature Correlation Analysis (NN-FCA) in two stages. Firstly, the feature selection step ranks features based on their importance in forecasting CHD risk. Subsequently, the feature correlation analysis stage explores correlations between features and the data output of each classifier. The evaluation conducted on a Korean dataset with 4146 individuals, where 3031 records had low risk and 1115 records had high risk of CHD, demonstrated the superiority of the proposed model AUROC curve: 0.749, over the FRS 0.393. In conclusion, NN-FCA, leveraging feature correlation analysis, outperforms Framingham risk score in CHD risk prediction, exhibiting a larger ROC curve and greater accuracy in predicting CHD risk within the Korean population. See the summary in Table 2.2.

Table 2.2: Summary of Deep Learning Models

Ref	Method	Dataset	Accuracy %
[60]	Higher order IoT wearable medical device to forecast heart illness with Boltzmann model (HOBDBN)	Hungarian	99.0
[61]	Proposed NN of CVD prediction	Heart Disease dataset	93.8
[62]	Neural Network in SAS Enterprise Miner Software	Cleveland	85.2
[63]	ANN model to assess the prognosis of carotid artery stenting	Cleveland	82.5
[65]	Codon-based backpropagation neural network method, with SVM, BPNN	Hungarian	90.0
[66]	Multi-Layer Perceptron for prediction	statLog	95.6
[67]	Multi-Layer Perceptron for prediction	Cleveland	84.5
[68]	Novel DL approach utilizing a CNN model	MIT-BIH and INCART datasets	98.4
[69]	Proposed CNN for CVD prediction and also LR, SVM, RF, ADA	NHANES	79.8
[70]	Using Big Data and XGB, and KNN to Predict Heart Disease	hypertensive patients in Shenzhen, China	94.3
[71]	Proposed Neural Network Model	Korean dataset	74.9
[72]	Improved LightGBM Model for CVD prediction	Framingham	93.0
[73]	CVD prediction using ML algorithms like KNN, and LR	CleveLand	87.0
[74]	Feature fusion-based healthcare monitoring system using SVM, LR, RF, DT, and NB	Health Care Big data	98.5
[75]	MDCNN Classifier Framework	Cleveland	98.2

Ref	Method	Dataset	Accuracy %
[76]	IoT based hybrid recommender system using SVM, NB, MLP, RF	100 cardiac patients' dataset	98.0

Menzies et al [77] introduce a unique approach to addressing the challenge of creating Software Effort Estimation (SEE) models, treating it as a multi-objective problem that explicitly and concurrently considers various performance measures. Utilizing a Multi-Objective Evolutionary Algorithm (MOEA) enhances the understanding of these performance measures, leading to the development of SEE models exhibiting superior overall performance compared to those not explicitly incorporating these measures. According to the research study, the evaluation metrics Mean Magnitude of the Relative Error (MMRE) and Logarithmic Standard Deviation (LSD) act slightly in opposition to one another. This affects the model's selection based on the desired evaluation measures. The motivation for employing MOEAs lies in capacity to consider all performance measures simultaneously, resulting in the creation of diverse ensembles likely to enhance overall performance. The study demonstrates that a MOEA effectively builds models by explicitly combining many performance metrics; for datasets containing 60 projects or more, the Pareto ensemble of MLPs typically outperforms backpropagation MLPs. The research underscores the flexibility of MOEAs, allowing software managers to emphasize specific performance measures while maintaining a balanced approach. The Pareto ensemble emerges as a versatile trade-off, accommodating diverse managerial preferences. Comparative analyses highlight the utility of MOEAs for both single and multicompany datasets, particularly excelling in heterogeneous data sets by elevating models that might not typically rank first in terms of performance.

With an exclusive focus on heart patient healthcare, Tuli et al [78] unveiled the ground-breaking HealthFog, a fog-based smart healthcare system that combines DL and IoT to automatically diagnose heart diseases. HealthFog effectively handles cardiac patient data from a range of Internet of Things (IoT) devices by acting as a fog service and integrating DL with Edge computing devices for useful heart disease analysis. This study addresses the resource-intensive nature of high-accuracy DL models. It achieves this by employing state-of-the-art model distribution and communication techniques like as ensembling, and by integrating complex networks into Edge computing concepts. Real-time analysis of cardiac patient data, neural network training on well-known datasets, and the implementation of a workable system that delivers immediate prediction results are all steps in the validation process. The efficacy of HealthFog is thoroughly assessed in a fog computing environment using the FogBus framework, accounting for factors including power consumption, network bandwidth, latency, jitter, training accuracy, testing accuracy, and execution time. Subsequent efforts will expand HealthFog for cost-effective implementation, taking into account different Quality of Service (QoS) attributes and fog-cloud pricing structures. Strongness and generality in the suggested architecture could enable its application to a wide range of fog computing applications, including smart city initiatives, traffic control, healthcare, and agriculture. The reach of HealthFog can be expanded to include other critical

healthcare domains including hepatitis, diabetes, and cancer, offering patients in these areas' effective services.

When compared to established predictive models (TIMI, MAGGIC, GRACE, and GWTG-HF scores) and alternative ML techniques (LR and RF), Kwon et al [79] demonstrate the superior predictive capabilities of a DL model based on ECG in forecasting in-hospital mortality among heart disease patients. The heightened performance of DL stems from its ability to intricately evaluate variable relationships and autonomously extract predictive features through multiple layers, surpassing the capabilities of traditional LR and RF models. It is crucial to acknowledge that DL and ML models, being devoid of medical knowledge-based rules, operate contextually, memorizing the characteristics of the derivation data. Authors use subgroup analysis and external validation to guarantee DL's robustness in a variety of scenarios. The study highlights the shortcomings of using AUROC to assess data that is unbalanced and promotes the use of Area Under the Precision-Recall Curve (AUPRC), particularly in situations where unusual occurrences occur infrequently, such as in-hospital mortality. Recognizing the significance of imbalanced data in model derivation, the study employs data processing methods to enhance the accuracy of the DL model, a challenge commonly encountered in medical data and clinical settings where non-event cases are predominant. Therefore, for researchers pursuing ML or DL investigations in the medical domain, comprehending AUPRC and implementing appropriate data processing techniques becomes crucial.

Table 2.3: List of Datasets from Literature Review

S/N	Dataset Name	Features	Records
1	StatLog [80]	13	270
2	Cleveland [81]	14	303
3	Framingham [82]	16	4,240
4	NHANES [83]	51	37,079
5	Cardiovascular disease dataset [84]	12	70,000

2.3 XAI Related Studies

XAI, or "eXplainable Artificial Intelligence" refers to a collection of measures and techniques that enable people to appreciate and rely on the output and results produced by ML algorithms. To deploy XAI and avoid naively trusting AI, a company must fully understand the decision-making processes of AI with model monitoring and responsibility. It can facilitate human comprehension and explanation of neural networks, DL, and ML algorithms. Many times, ML models are perceived as unintelligible "black boxes" Some of the hardest neural networks for humans to comprehend are those used in DL. Biasness factor has always been a risk when training AI models, and it is often based on factors like region, age, gender, or race. Moreover, AI model performance may degrade or drift when training and production data differ. This means that a company needs to continuously monitor and maintain its models in order to promote AI explainability and assess the business impact of implementing such algorithms.

Adadi and Berrada [27] adopt a holistic approach, akin to the comprehensive assimilation of new topics, by addressing the Five W's and How (What, Who, When, Why, Where, and How) to encompass all facets of XAI. To map the expansive landscape of XAI research, the survey delves into a range of explainability approaches, offering a thorough examination from various perspectives. Discoveries underscore that XAI extends beyond the confines of a laboratory, influencing diverse application domains. The research also highlights how explainability methodologies now in use pay insufficient attention to the human element, and it exposes a lack of formality in problem formulation and precise definitions. Essentially, other interesting pathways of AI system explainability have gone mostly untapped due to the concentration on interpreting ML models. The culmination of this exploration signals the imperative for substantial future efforts to address challenges and unresolved issues within the realm of XAI. Additionally, explainable AI supports productive AI use, model auditability, and end-user trust. Moreover, it reduces the reputational, legal, security, and compliance concerns associated with production AI.

Clinical Decision Support Systems (CDSS) [85] are intended to support human decision-making by being reliable, simple to use, and helpful. Explainability is essential to reaching these objectives. Explainability enables engineers to spot flaws in a system and gives physicians peace of mind while using CDSS support to make judgments. The authors of this evaluation of XAI in CDSS concentrated on the "where" and "how" of XAI use in CDSS, and they were able to assess some of the benefits that had been obtained as well as pinpoint future needs in this field. The selection of techniques for effectively and informatively presenting explanations continues to be a major difficulty. There is still a lot of effort to be done to incorporate helpful explainability into CDSS. To thoroughly prove how explainability may be applied in this significant setting, studies concentrating on all stages of CDSS development are needed.

S. Das et al. [86] concentrate on using XAI to reduce dimensionality without compromising the classification accuracy of heart disease. Using SHAP, four explainable ML models represented the feature weights (FW), feature contributions (FC), and for every CFV feature in order to obtain the desired outcomes. The compact dimensional feature subset (FS) was obtained for FC and FW.

V. Belle and I. Papantonis [87], the authors go into greater detail on the XAI model. It states that ML models are rapidly being used in a wide variety of industries. However, because of the increasing occurrence and complexity of methodologies, business shareholders are becoming progressively worried about trained AI model disadvantages, biases with training data subset, and so on. Similarly, data science practitioners are frequently unaware of methodologies emerging from academic literature or may struggle to comprehend the differences between different methods, thus they resort to industry norms. Visualizing the black box can also be supported by the user interface module's interpretability design [88]. Transmission, discourse, experience, optimal behavior, control, tool use, and embodied action are examples of interaction variables that are crucial to consider while building an AI-based system. Four principles of human-centered design for ML improve human user comprehension by employing several explanation-generating strategies.

American's DARPA (Defense Advanced Research Projects Agency) [89] researched interpretable AI technology in 2019. This study told us certain indicators can be used to assess the success of these explainable, interpretable models.

i) User Gratification

- Simplicity of the interpretability
- Usefulness of the explanation

ii) Psychological Model

- Sympathetic individual choices
- Considerate the general model
- Strength/softness calculation
- What and How Questions in prediction

iii) Duty Performance

- Does the interpretation help the user make better decisions and perform better on jobs?
- Artificial decision jobs were announced to identify the user's kind

iv) Trust Calculation

- Suitable future use and faith on the system

In order to investigate how AI systems ought to inform end users of their decisions, Laato et al [90] carried out a thorough examination of the literature. Five high-level objectives for AI system communication were determined by synthesizing the literature: understandability, trustworthiness, transparency, controllability, and fairness. Design suggestions were put forth, stressing customized and on-demand explanations and concentrating on essential features as opposed to the system as a whole. The study accepts that there are trade-offs in the explanations of AI systems and that there isn't a perfect answer. In order to improve understandability, fairness, trustworthiness, controllability, and transparency of AI systems for end users, a design framework was created. This framework contributes to AI governance. Three major contributions emerged from the systematic literature study, which included examining twenty-five empirical research articles: establishing communication objectives, gathering and creating design recommendations, and presenting a combined design framework. AI system communication designers and XAI professionals can benefit greatly from this approach, which facilitates user-oriented communication in line with AI governance objectives.

Guleria et al [91] delve at the benefits, drawbacks, and contributions of AI and ML in healthcare, highlighting an experimental strategy that employs ML approaches to forecast cardiac disease. The SVM algorithm demonstrates superior performance with an 82.5% accuracy in heart disease classification. Various ML algorithms, including AdaBoost, bagged trees, Gaussian NB, SVM, KNN, and LR are explored, along with XAI techniques for proper interpretability. The study recognizes limits in dataset scope and size, advocating for self-learning models with minimal data requirements. The research study highlights the status of interpretability and trustworthiness in decision-making process of models, supporting for ensemble

classification models within the XAI framework. The evaluation of these XAI models using several metrics exposes the strong performance of XAI-driven of ML algorithms like SVM, LR, and NB. These ML algorithms achieve an accuracy of 89%, creating them as compelling replacements in comparison to the already implemented models.

In this study [92], the effort lies on the proportional analysis of multiple ML algorithms for forecasting heart attack rates based on several features. The research study delves into defining the position of these factors and evaluating the prediction accuracy of the algorithms. Remarkably, the XGB Classifier appears as a better performer, achieving a noteworthy accuracy rate of 86.885%. The balancing model complexity, explainability, and prediction performance is essential for the AI scientists. Complex algorithms can make it hard to appreciate how models work, known as the "black box" problem. To address this, XAI methods like Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) are useful. SHAP visually clarifies model's decision-making process and shows feature rank, as applied to the KNN algorithm in this study. LIME assists to explain how the KNN model works by providing rich insights into its predicting calculation process. This research study not only advances heart attack forecast using different ML algorithms but also highlights the efficiency of SHAP and LIME in explaining the performance of the KNN algorithm.

In this research study [93], Nascita et al attached the power of XAI to clarify, enhance, and implement multi-modal DL approaches for addressing several Traffic Classification (TC) responsibilities. The study focused on designing, implementing, and evaluating an advanced multi-modal multi-task DL based traffic classifier. It was enhanced in phases, using algorithms of XAI to guide the process. This iterative process culminated in the development of the DISTILLER-EVOLVED model, a result of successive refinement within the overarching DISTILLER framework. The initial approach (DISTILLER-EMBEDDINGS) outperformed multiple baselines multitask DL classifiers, including the prior state of the art, according to evaluation on the ISCX VPNNONVPN dataset, which was annotated for three TC tasks (encapsulation, traffic type, and application recognition). The authors emphasized the importance of payload information even in the face of a significant amount of encrypted communication by providing a thorough explanation of the modality contributions to each task through the use of interpretability techniques like Deep SHAP and Integrated Gradients. In comparison to DISTILLER-ORIGINAL, DISTILLER-EMBEDDINGS showed a more equitable relevance between input modalities. The authors presented an improved version, DISTILLER-EARLIER, whose dependability was tested by calibration, building on interpretability insights. The DISTILLER-CALIBRATED classifier maintained performance and greatly increased reliability by utilizing label smoothing. Pruning was found to be the best compression strategy after model size reduction techniques were investigated. This resulted in the creation of DISTILLER-EVOLVED, which outperformed DISTILLER-ORIGINAL in terms of performance, interpretability, reliability, and memory efficiency. See the Table 2.4.

Table 2.4: Summary of XAI Literature Review

Ref	Method	Outcome
[27]	XAI techniques on ML models	Five W's and How (What, Who, When, why, Where, and How)
[85]	Clinical Decision Support Systems (CDSS)	Focused on the "where" and "how" of XAI
[86]	eXplainable AI on ML techniques like DT, RF, and SVM	Identifies the top features
[87]	ML model Local and Global explanation	Feature weights
[88]	Interpretability design of the user interface module	Visualizing the black box
[89]	DARPA (Defense Advanced Research Projects Agency) research on interpretable AI Technology in 2019	User Gratification, Psychological Model, Trust Calculation, etc
[90]	Explore how AI systems should communicate their decisions to end users	Understandability, Trustworthiness, Transparency, Controllability, Fairness
[91]	Importance of interpretability and trustworthiness in decision models, advocating for ensemble classification models	SVM, KNN, AdaBoost, LR, and Gaussian NB with XAI
[92]	XAI methods SHAP and LIME with XGBoost and KNN model	Model complexity, explainability, and prediction performance
[93]	Traffic Classification, developing DISTILLER-EVOLVED model	Model performance, interpretability, reliability, and memory efficiency

Chapter 3: Proposed Research Methodology

In this chapter, this study discusses the proposed methodology and pictorial representation of it. A list of challenges in the selected dataset is also given in this chapter. In addition, the details of the feature values and their ambiguousness are also discussed.

3.1 Dataset Exploration

This study will use NHANES, the selected dataset has unique features, although some of them are common with other datasets from previous studies. First, datasets are preprocessed individually, deleting missing data, removing duplicates, converting category values to numerical values, and so on Data Exploration

Within the realm of cardiovascular health investigation, the dataset under consideration emerges as an expansive and intricate repository designed for the precise prediction of CHD. Structured in the XLS file format, this dataset encompasses a substantial 37,079 individual records, each meticulously capturing a unique amalgamation of biological and demographic elements. Comprising a rich tapestry of 51 distinct features see in Table 3.1, the dataset provides an intricate panorama of physiological and socioeconomic dimensions.

Table 3.1:List of Features in Dataset

S/N	Features	S/N	Features	S/N	Features	S/N	Features
1	SEQN	14	Monocyte	27	Albumin	40	Uric.Acid
2	Gender	15	Eosinophils	28	ALP	41	Triglycerides
3	Age	16	Basophils	29	AST	42	Total-Cholesterol
4	Annual-Family-Income	17	Red-Blood-Cells	30	ALT	43	HDL
5	Ratio-Family-Income-Poverty	18	Hemoglobin	31	Cholesterol	44	Glycohemoglobin
6	X60-sec-pulse	19	Mean-Cell-Vol	32	Creatinine	45	Vigorous-work
7	Systolic	20	Mean-Cell-Hgb-Conc.	33	Glucose	46	Moderate-work
8	Diastolic	21	Mean-cell-Hemoglobin	34	GGT	47	Health-Insurance
9	Weight	22	Platelet-count	35	Iron	48	Diabetes
10	Height	23	Mean-Platelet-Vol	36	LDH	49	Blood-Rel-Diabetes
11	Body-Mass-Index	24	Segmented-Neutrophils	37	Phosphorus	50	Blood-Rel-Stroke
12	White-Blood-Cells	25	Hematocrit	38	Bilirubin	51	CoronaryHeartDisease (Target Class)
13	Lymphocyte	26	Red-Cell-Distribution-Width	39	Protein		

From fundamental demographic indicators such as “Gender” and “Age” to economic metrics like “Annual-Family-Income” and “Ratio-Family-Income-Poverty” and critical physiological markers such as “Systolic” and “Diastolic” blood pressure readings, the dataset offers a comprehensive insight into the subject's profile.

The addition of hematological data features such as “White-Blood-Cells”, “Hemoglobin” and “Platelet-count” increases a layer of hardness, while metabolic data features like “Cholesterol”, “Glucose” and “Triglycerides” contribute to a nuanced understanding of the biochemical system. The selected dataset centers on a key target called “CoronaryHeartDisease”. The main outcome of the model is to predict the target class correctly. All other data features help in adjusting the weights of neurons in the model. The focus on Coronary Heart Disease allows the model to identify patterns and feature values linked to it. For creating an accurate prediction model, the model predicts the target class with high accuracy, precision, and recall.

3.1.1 Correlation Matrix and Heat Map

The dataset captures the complex associations between many features which are directly related to the subject disease. It is a valuable source for understanding how this disease develops. By exploring these connections between features, the research study finds important patterns and dependencies concerning the subject disease. A correlation matrix is plotted using Python Pandas and Matplotlib libraries, shown in Figure 3.1, helps to disclose these relationships and dependencies.

This correlation matrix displays 30 diverse features and captures the pairwise correlations between each feature. The linear association of these feature values is plotted within the range of -1 to +1. The strong positive association is implied by a number near +1, and the strong negative association is indicated by a number near -1.

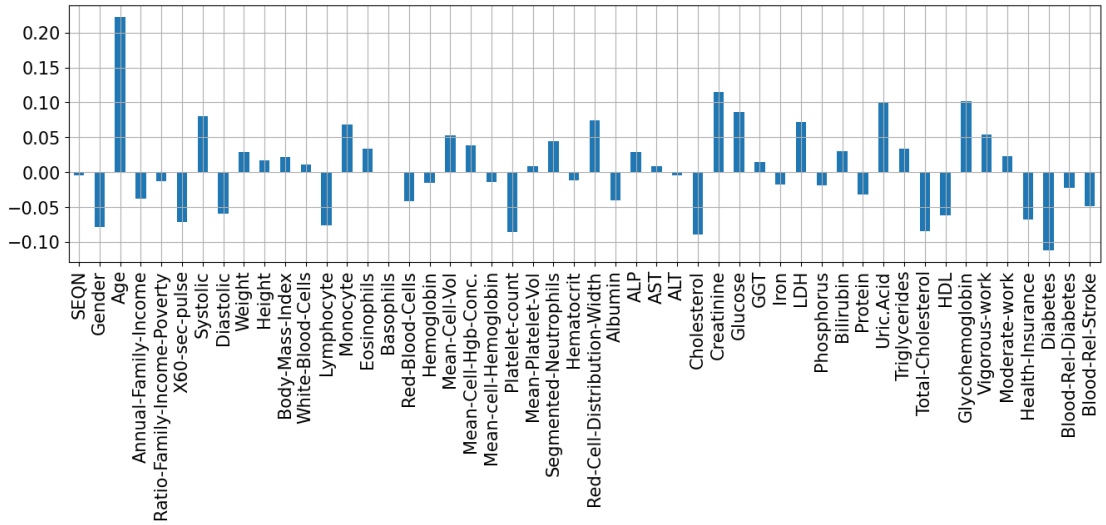

Figure 3.1: Correlation Matrix of Features

A heat-map of feature correlation is displayed in Figure 3.2. Insights from this plot help researchers recognize the detailed correlation and dependences with color visualization. The plot guides in choosing the most vital features for investigation. This also helps in less computation time and resources required to train the model.

3.1.2 Gender Distribution

The dataset has an approximately equal gender distribution among its 37,079 records, with 51% male and 49% female. This balance of samples with respect to gender is important for reasonable and inclusive analysis of heart disease. A gender-balanced dataset helps to train such a model that predicts more accurately and reliably.

This equal representation allows researchers to study connections between physical and social factors effectively. It confirms that analyses and models mirror both genders fairly.

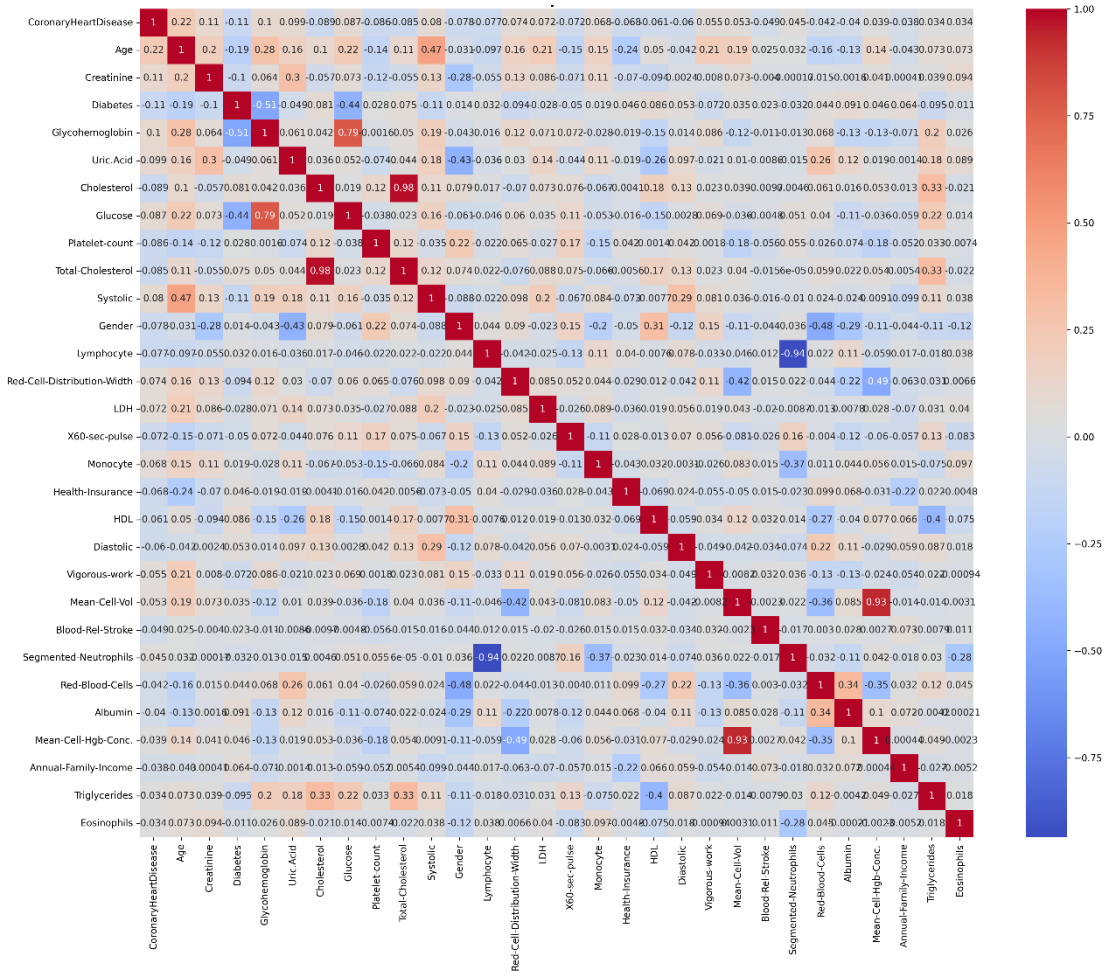

Figure 3.2: Heatmap of Features

3.1.3 Age-Base Distribution

The research study observes age distribution across all records of the dataset, grouping individuals into different age ranges (shown in Figure 3.3). A key focus is on ages 30 to 75, which shows a wide span of adult life, with a density of 0.03% having heart disease. The density rises to 0.07% for the age group of 75 to 88 years old, reflecting more detailed data for older adults. This age grouping helps in considerate health patterns across diverse phases of life. This style of research is essential for finding age-related tendencies and building good, reliable, and accurate predictive models.

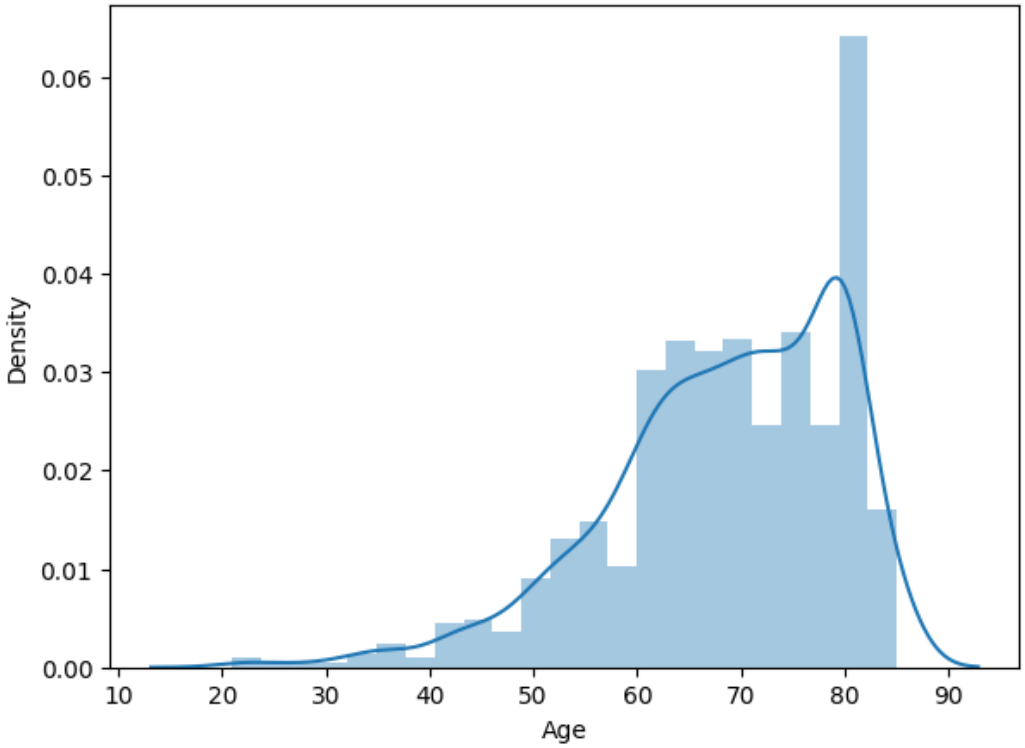

Figure 3.3: Age-based distribution of Dataset

3.1.4 Target Class Distribution and Data Imbalance

The research study explores the dataset and reveals a notable class imbalance, with 35,571 records for healthy individuals and 1,508 for those with heart disease. This results in 95.93% healthy samples and 4.07% indicating heart disease, see in Table 3.2. Such imbalance can generate tests during model training, as algorithms may biased toward predicting the majority class (healthy). To overcome this problem, the research study uses "imblearn" Python library's method named "SMOTE". This method creates synthetic samples of the minority class (heart disease patients). This process balances the data and promotes fair representation of both target classes.

Table 3.2: Dataset Imbalanced Statistics

Total Records		37,079	
Index	**Feature Name**	**Class in Features**	**Number of Records**
1	Target Class	Healthy	35,571
		Patient	1,508

3.2 Preprocessing

The pre-processing pipeline for this dataset follows the systematic phases using Python programming language. First, the dataset XLS file is read using the Pandas library, and its structure is observed, including shape, column names, and record count. This is an important step to understand the size and list of features of the dataset. The research study observed that there are missing values in the dataset, which need to be removed those samples for the quality of data feed to the model. Duplicate records are also analyzed and removed to further enhance data quality. Finally, the shape of the dataset is checked to confirm these changes. This makes

the dataset ready for future tasks. Using Python and the Pandas library shows a clear and careful way to prepare the dataset. This approach is important for any type of data science research and AI model development.

3.3 Data Balancing and Augmentation

The systematic methods SMOTE and ADASYN techniques from Python "imblearn" library are used to overcome the class imbalance in the dataset and generate two datasets. The study conducts two experiments on these datasets, see more details in Table 3.3. The first experiment uses the SMOTE augmented dataset, while the second experiment uses the ADASYN augmented dataset using the same MLP model architecture. The scaling of the feature values using Python Science-Kit-Learn (sklean) preprocessing method named "StandardScaler" further standardizes the augmented data. This balanced class and standard-scaled dataset help the proposed MLP model to learn patterns of feature values and improve performance in identifying the target class. Next, the dataset is divided into training and testing subsets using the train_test_split method, with 80% for training and 20% for testing. This split saves the target class distribution integral.

Table 3.3: Balanced Dataset

Index	Feature Name	Total Records	Class Name	Records
1	SMOTE Technique	71,142	Healthy	35,571
			Patient	35,571
2	ADASYN Technique	70,699	Healthy	35,571
			Patient	35,128

3.4 Splitting data to Subsets

After all these pre-processing steps, split the dataset into train and test segments with 80% and 20% ratios, respectively. The major goal is to put the attention model to the test on this dataset and see evaluation measures. After the training section, the proposed method uses the testing subset of data for the evaluation of the newly trained model. This study implements the XAI techniques SHAP on this trained attention model to see the inner workings, and visual representation of feature importance, for the class prediction.

3.5 Architecture of Attention Base Multi-Layer Perceptron Model

The architecture of the presented MLP Model with Attention layers unfolds in a sequence of interconnected layers designed for a specific computational task. The proposed DL model is designed to enhance the predictive capabilities for coronary heart disease seen in Figure 3.4. Comprising various layers, this model begins with an input layer of 40 dimensions. The subsequent layers include densely connected layers, each contributing to the extraction and transformation of essential features. Specifically, a dense layer with 80 neurons is employed, followed by three additional dense layers with 40 neurons each. This output undergoes further processing through an additional dense layer with 20 neurons. The final output layer only contains 1 neuron because we only want to predict yes or no result about the subject disease. The overall model structure is meticulously fine-tuned, with a total of 9,615 parameters. This updated architecture is poised to elevate the model's predictive accuracy and robustness in cardiovascular health prediction.

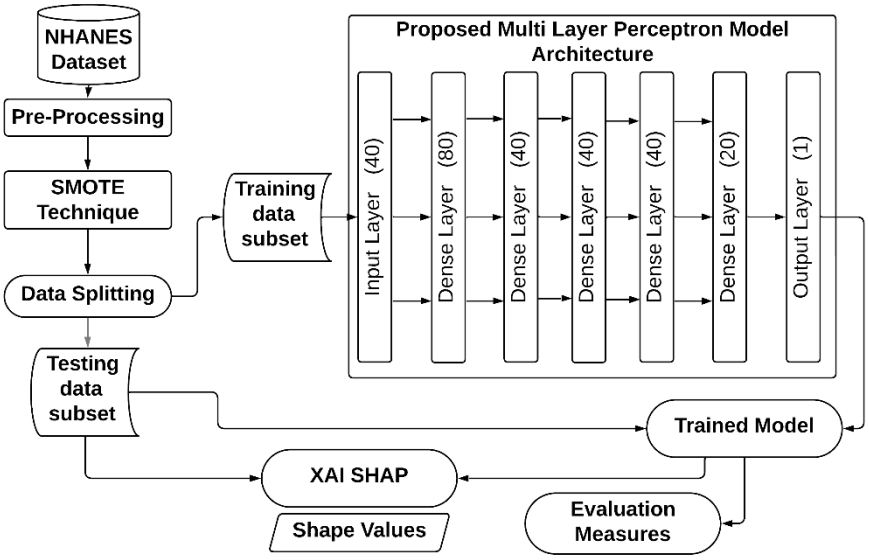

Figure 3.4: Proposed MLP and Attention Layer Framework

The comprehensive design of this model, integrating attention mechanisms and batch normalization, exemplifies a sophisticated approach to feature extraction and learning. The utilization of these techniques aims to enhance the interpretability and predictive performance of the model in the context of subject disease classification.

3.6 Mathematical Calculation of Proposed Model

Let X represent the input features, and W and b denote the weight and bias parameters for each layer. The activation function is denoted as f, and α represents the attention weights.

- Dense Layer (input 40 features)

$$Z_1 = X.W_1 + b_1$$
$$A_1 = f(Z_1)$$

- Dense Layer (expends to 80 nodes)

$$Z_2 = A_1.W_2 + b_2$$
$$A_2 = f(Z_2)$$

- Dense Layer (reduce to 40 features)

$$Z_3 = A_2.W_3 + b_3$$
$$A_3 = f(Z_3)$$

- Dense Layer (40 features)

$$Z_4 = A_3.W_4 + b_4$$
$$A_4 = f(Z_4)$$

- Dense Layer (40 features)

$$Z_5 = A_4.W_5 + b_5$$
$$A_5 = f(Z_5)$$

- Dense Layer (reduce to 25 nodes)

$$Z_6 = A_5.W_6 + b_6$$
$$A_6 = f(Z_6)$$

- Output Layer (only 1 node)

$$Final\_Output = A_6.W_{output} + b_{output}$$

3.7 Performance Measures

The assessment of model performance is paramount to gauge its efficacy and reliability. Various evaluation measures provide insights into different aspects of a model's behavior.

3.7.1 Accuracy

A basic indicator that is determined by dividing the total number of instances by the ratio of correctly predicted instances (True Positives and True Negatives). Although it offers a general indicator of accuracy and efficacy, it might not be adequate in situations when there is an uneven distribution of classes.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3.7.2 Precision

The accuracy of optimistic forecasts is the main emphasis of precision. The ratio of True Positives to the total of True Positives and False Positives is used to compute it. In medical diagnosis applications, a low rate of false-positives rate is reflected to a high precision score.

$$\Pr ecision = \frac{TP}{TP + FP}$$

3.7.3 Sensitivity or Recall

Sensitivity or recall gives insights into how well the model can differentiate true positive cases from false positive instances. This evaluation measure is also called the True Positive Rate. It is the ratio of True Positives to the total number of True Positives and False Negatives. This is critical in situations where missing positive prediction is highly undesirable, like this research study.

$$Sensitivity = \frac{TP}{TP + FN}$$

3.7.4 AUC Score and AUROC Curve

One important evaluation metric is the AUC Score, commonly used in binary classification problem-solving. It is a visual plot between the true positive rate against the false positive rate. The plot shows the area under the ROC curve. A higher AUC score directs a better ability to differentiate between classes. The AUROC curve (Area Under the Receiver Operating Characteristic) visually signifies the performance of the model at diverse classification thresholds.

3.7.5 F1 Score

The F1 Score is another famous evaluation metric that balances recall and precision. It provides a stronger view of how well a model accomplishes the correct prediction task, especially with unbalanced datasets. The basically F1 Score is the harmonic mean of precision and recall.

$$F1\_Score = \frac{2}{\frac{1}{Sensitivity} + \frac{1}{\Pr ecision}}$$

In summary, the specific goals and occurrence of the problem help to decide which evaluation metrics to use.

Chapter 4: Results and Discussions

In this chapter, this study discusses the preprocessing steps and making imbalanced data into a balanced dataset through data augmentation. The proposed model architecture is also discussed in detail. At the end of this chapter, the study shows the evaluation measures of the proposed model and show how capable is this system to implement in real-world situation.

4.1 Experimental Setup

In the pursuit of developing an advanced predictive model for subject disease, a comprehensive experimental setup was established. The model's efficiency and resilience were greatly enhanced by the hardware and software environment.

4.1.1 Hardware Configuration

The research is carried out using a Lenovo laptop computer system with an Intel(R) Core (TM) i5 10^{th} generation CPU @ 2.50 GHz and a clock speed of 2.50 GHz. There was 16.0 GB of installed RAM on the machine, of which 15.9 GB could be used. The operating system, Windows 11 Pro for Workstations Version 23H2, provided a 64-bit environment on an x64-based processor architecture. The CUDA-enabled NVIDIA GeForce GTX 1650 with a total memory of 4.29 GB played a pivotal role in accelerating computations.

4.1.2 Software Environment

The required software includes Visual Studio Code (VScode), version 1.85.1, serves as the main programming or coding environment (IDE). The most famous programming language for building ML pipelines, the Python (version 3.10) used for the experiment. Key libraries included Pandas (2.1.2), NumPy (1.24.3), Matplotlib (3.8.1), Seaborn (0.13.0), Scikit-Learn (1.3.2), imbalanced-learn (imbLearn) with SMOTE (0.11.0), TensorFlow (2.15.0), Keras (2.15.0), and SHAP (0.44.0). These tools simplified data manipulation, analysis, and the model development (training and evaluation).

4.1.3 CUDA and GPU Configuration

The computing system had one CUDA device, an NVIDIA GeForce GTX 1650, with 4.29 GB of memory. The CUDA library, version 8700 is installed to utilize the parallel processing power during model training.

4.1.4 Development Libraries and Frameworks

The proposed model is developed using TensorFlow, with version 2.15. This TensorFlow framework provided an understandable architecture for building and developing MLP models. The Python “imblean” library method named SMOTE is used to overcome the class imbalance problem in the dataset.

This experimentation setup is used to explore the dataset, perform pre-processing steps on it and generate synthetic minority class samples for class balancing, design the architecture of the proposed model, train and evaluate the model, and at the last use XAI, SHAP algorithms to explore the model prediction behavior and feature ranking.

4.2 Results and Discussion

The proposed MLP model architecture on the balanced dataset shows exemplary performance score on essential evaluation metrics, like accuracy, recall, precision, etc. The results are discussed below in detail.

4.2.1 Imbalanced Dataset Results

The dataset under consideration exhibits a substantial class imbalance, with a test dataset containing 80% records of total of 71,142 records. The model underwent training for 40 epochs and gives a training accuracy of more than 98% see in Figure 4.1.

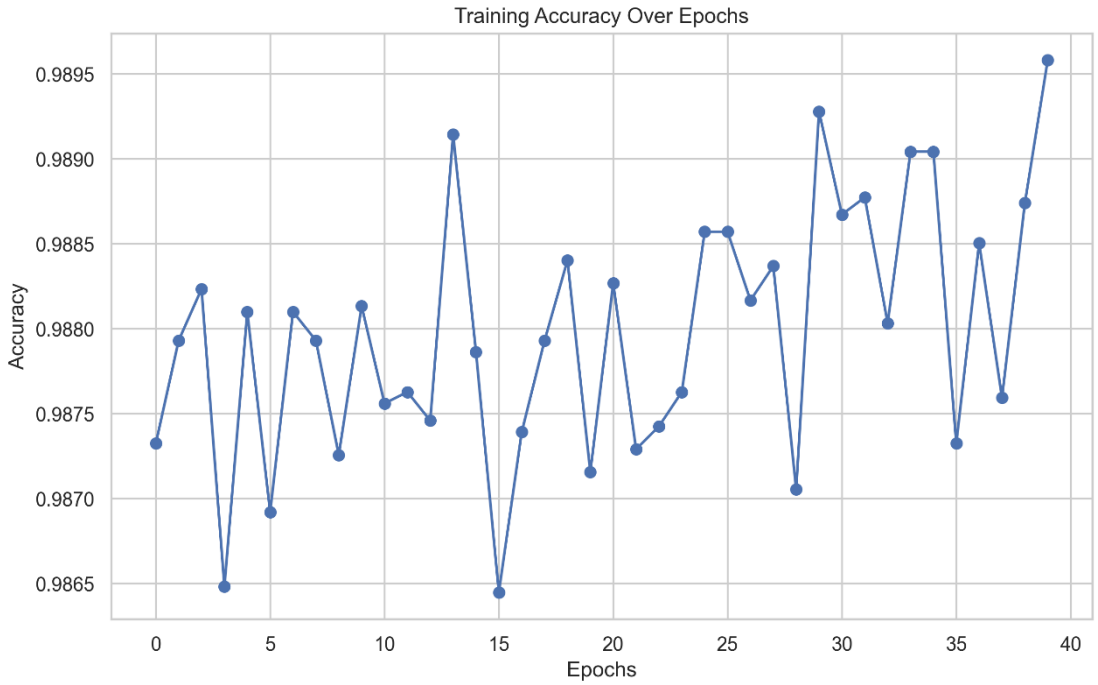

Figure 4.1: Training Accuracy on Imbalanced Dataset

The model is evaluated on the testing dataset and give the high accuracy but low other evaluation measures. The confusion matrix shows that the number of healthy patient records is much greater than the number of heart disease patient in the testing dataset seen in Figure 4.2. This shows that the model is biased toward the healthy class.

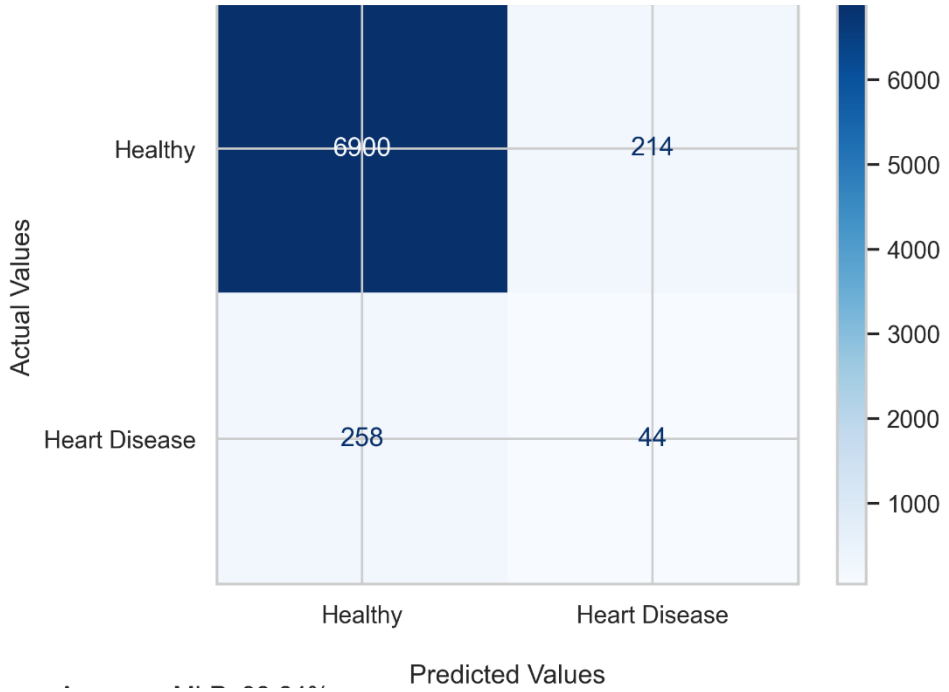

Figure 4.2: Confusion Matrix on Imbalanced Dataset

This imbalance is further emphasized by the precision-recall trade-off, where achieving high precision comes at the cost of lower recall. The relatively high precision of 69.23% suggests that when the model outputs the positive class, it is often correct, but the low recall seen in Table 4.1 indicates a substantial number of false negatives. This discrepancy is critical in applications where correctly predicting the positive instances is of utmost importance, like in medical diagnoses or fraud detection.

Table 4.1: Evaluation Measures on Imbalanced Dataset

S/N	Measure Nomenclature	Percentage
1	Accuracy	96.00
2	Recall	2.98
3	Precision	69.23
4	F1 Score	7.83
5	AUC Score	90.03

The F1 score, a measure that strikes a compromise between recall and precision, captures the subtleties of the model's performance and is notably low at 7.83%. The model's capability to correctly predict between the two classes is verified by the AUC score which is 90.03%. Handling the problem generated from rough data distribution is crucial. Overcoming this class imbalance problem is essential to improve model performance on such data. Analyzing specific cases where the model misclassifies positive instances can expose areas for enhancement. The high accuracy score is important, but simply focusing on it is not enough for class-imbalanced datasets. When missing positive cases has a high cost, balancing recall and precision value is also important, as well as understanding the real-world effect of false negatives classification.

4.2.2 Balanced Dataset Results

The proposed model performs very outstanding and gives high evaluation measure values, see the confusion matrix in Figure 4.3. The model test accuracy is 97.10%, which shows it can reliably differentiate between heart patients and healthy persons. With a recall value of 97.85%, the model does a great job of classifying actual positive records in the dataset. The model testing precision value is 96.40%, reflecting the model's durable accuracy in cataloging cases as positive or negative correctly see in Table 4.2.

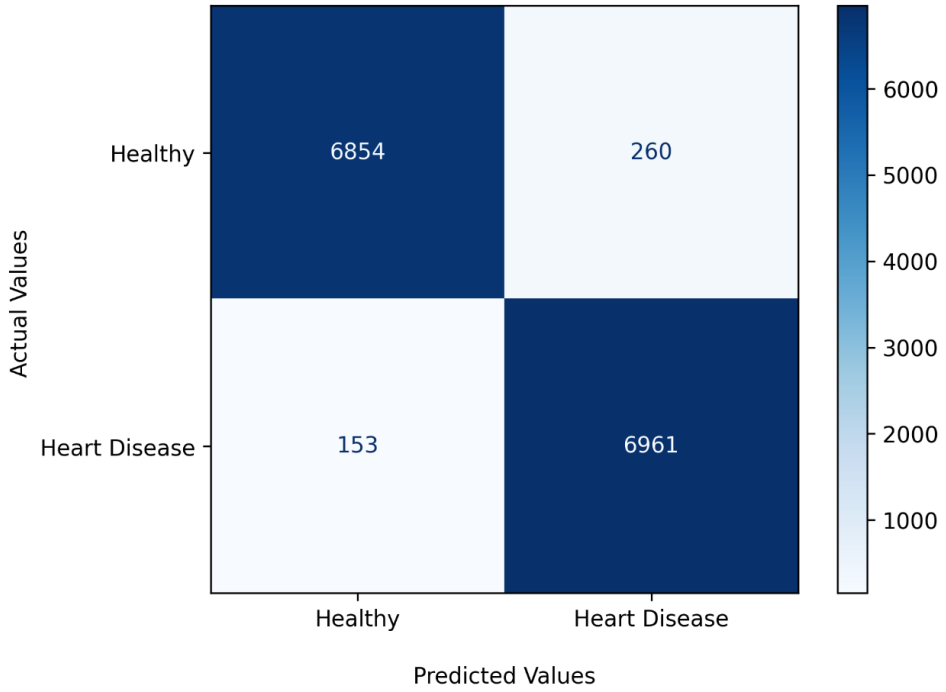

Figure 4.3: Confusion Matrix on Balanced Dataset

The F1-score is 66.67%, which echoes a balanced performance between precision and recall. The AUC value, which is used to measure the Receiver Operating Characteristic (ROC) curve, is important for evaluating how well the model can differentiate between target classes.

Table 4.2: Evaluation Measures on Balanced Dataset

S/N	Measure Nomenclature	Percentage
1	Accuracy	97.10
2	Recall	97.85
3	Precision	96.40
4	F1 Score	66.67
5	AUC Score	99.42

With a test AUC of 99.42%, the model does a great job of classifying positive cases separately from negative ones with high accuracy. This high AUC in shows the model's strong ability to separate different real-world scenarios.

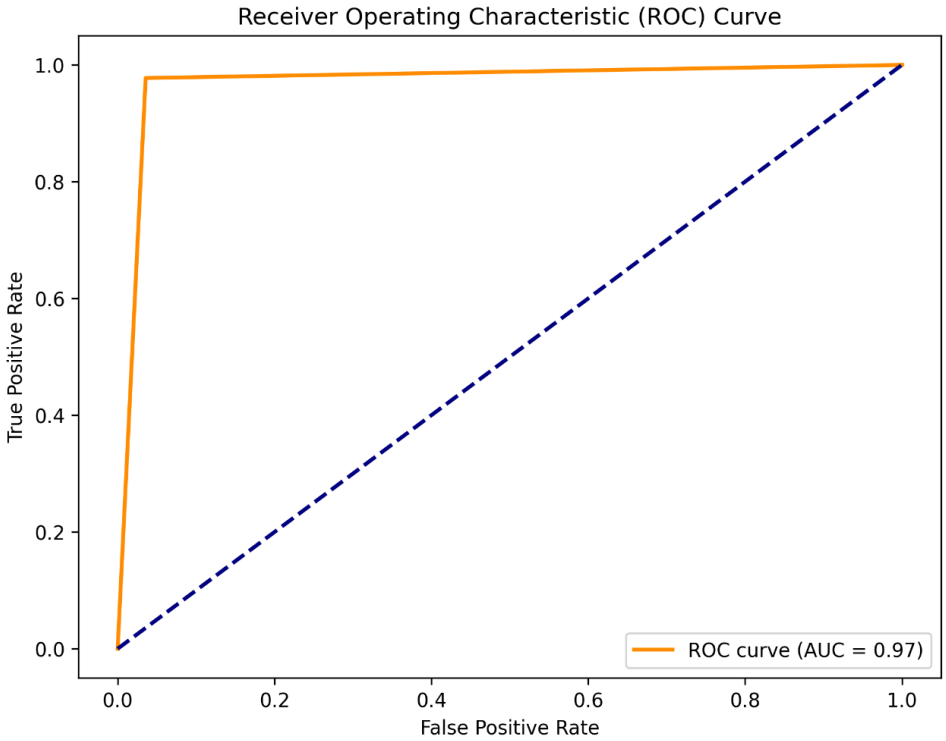

Figure 4.4: ROC Curve of Proposed Model

In comparison the both experiments, the study proves that the class imbalance problem generates a bad model. The model is biased toward the majority class and predicts many minority class samples to majority class samples. It is not suitable to deploy such type of model in real-world scenarios. In the other hand, the same model trained well on the balanced dataset and gave very good evaluation measures. The evaluation values are compared in Table 4.3. If this model is deployed in real-world scenarios, it performs well and gives accurate results similar to medical professional diagnoses.

Table 4.3: Comparison of Both Experiments

S/N	Measure	Imbalanced Data	Balanced Data
1	Accuracy	96.00	97.10
2	Recall	2.98	97.85
3	Precision	69.23	96.40
4	F1 Score	7.83	66.67
5	AUC Score	90.03	99.42

4.3 SHapley Additive exPlanations (SHAP) for Model Analysis

Shapley Additive exPlanations (SHAP) is currently a very famous method that helps to explain the prediction behavior of ML and neural network models. This model-agnostic technique is simple and can be applied to any model to help understand its behavior and improve evaluation measures. SHAP explanations classify key characters and give a very clear view of how each feature value impacts the final predicted result.

4.3.1 SHAP Plots (Imbalanced Dataset)

By experimenting of the framework on the imbalanced dataset and the associated evaluation measures, SHAP values can offer important insights into how the model calculates its weight matrix and comes to this final prediction. The model exhibits distinguished performance characteristics as designated by the above-mentioned evaluation measures including the confusion matrix. The confusion matrix exposes that while the testing accuracy is high (95.9951%), the recall measure value for the positive class is considerably very low (2.98%). This recall value indicates that the model has difficulty accurately classifying the minority class, and this could result in an increase in false negative classification. The precision value (69.23%) is quite high but it indicates that the proposed architecture is likely to be correct when it predicts the positive class only. Using the XAI SHAP algorithm at this stage enables the feature-by-feature interpretation of the decisions making process of the model. To see which features play an important role in predicting a positive or negative class, SHAP values measure the overall feature value effect on the result see Figure 4.5 and Figure 4.6. In the context of imbalanced datasets, both plots show feature importance is vital.

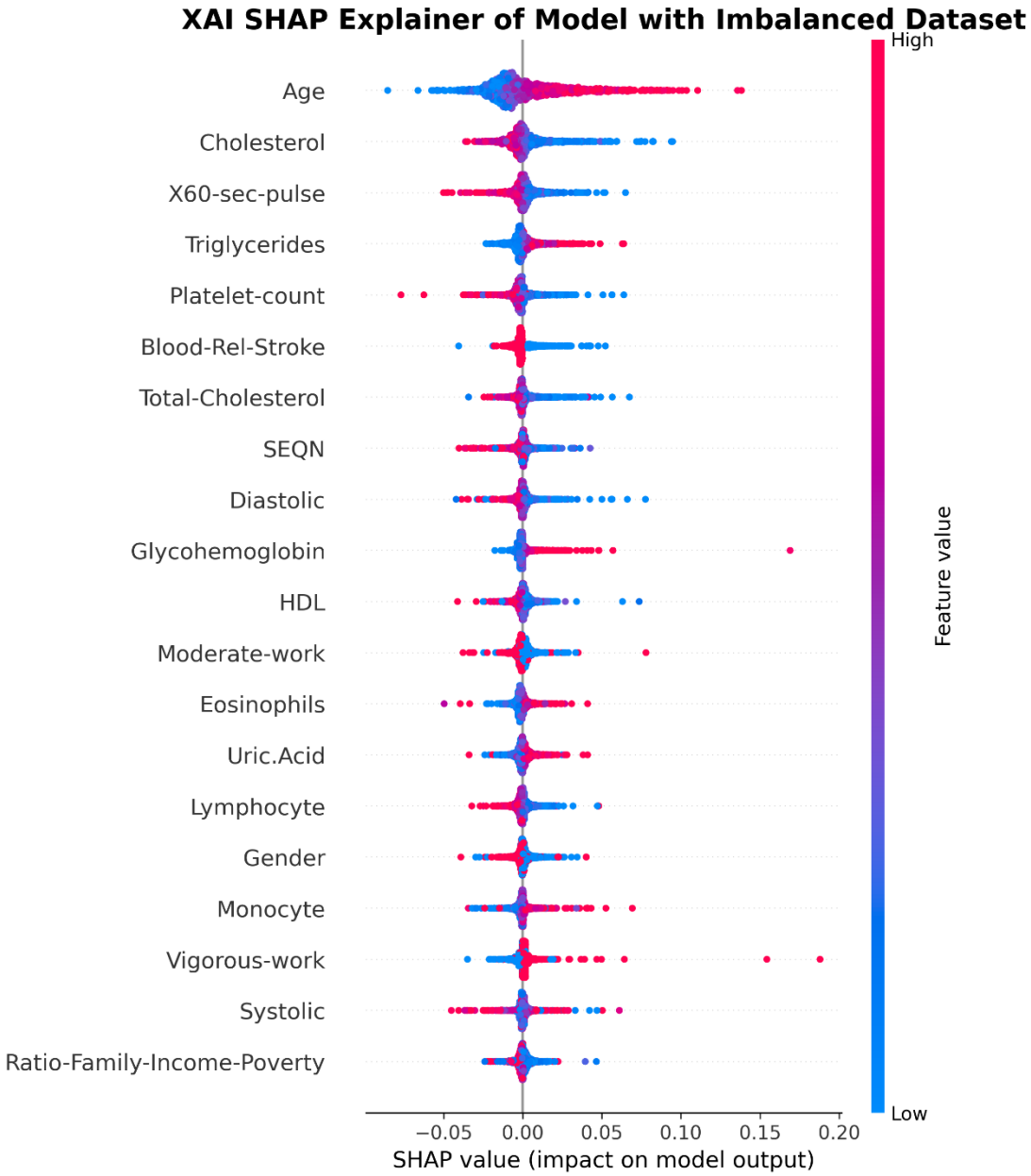

Figure 4.5: SHAP Feature Ranking Summary Plot (Imbalanced Dataset)

SHAP feature ranking analysis might disclose that crucial features play a greater role in contributing to false negative class. This analysis can guide additional model improvements, like feature engineering or targeted data augmentation for the minority class (heart patients records).

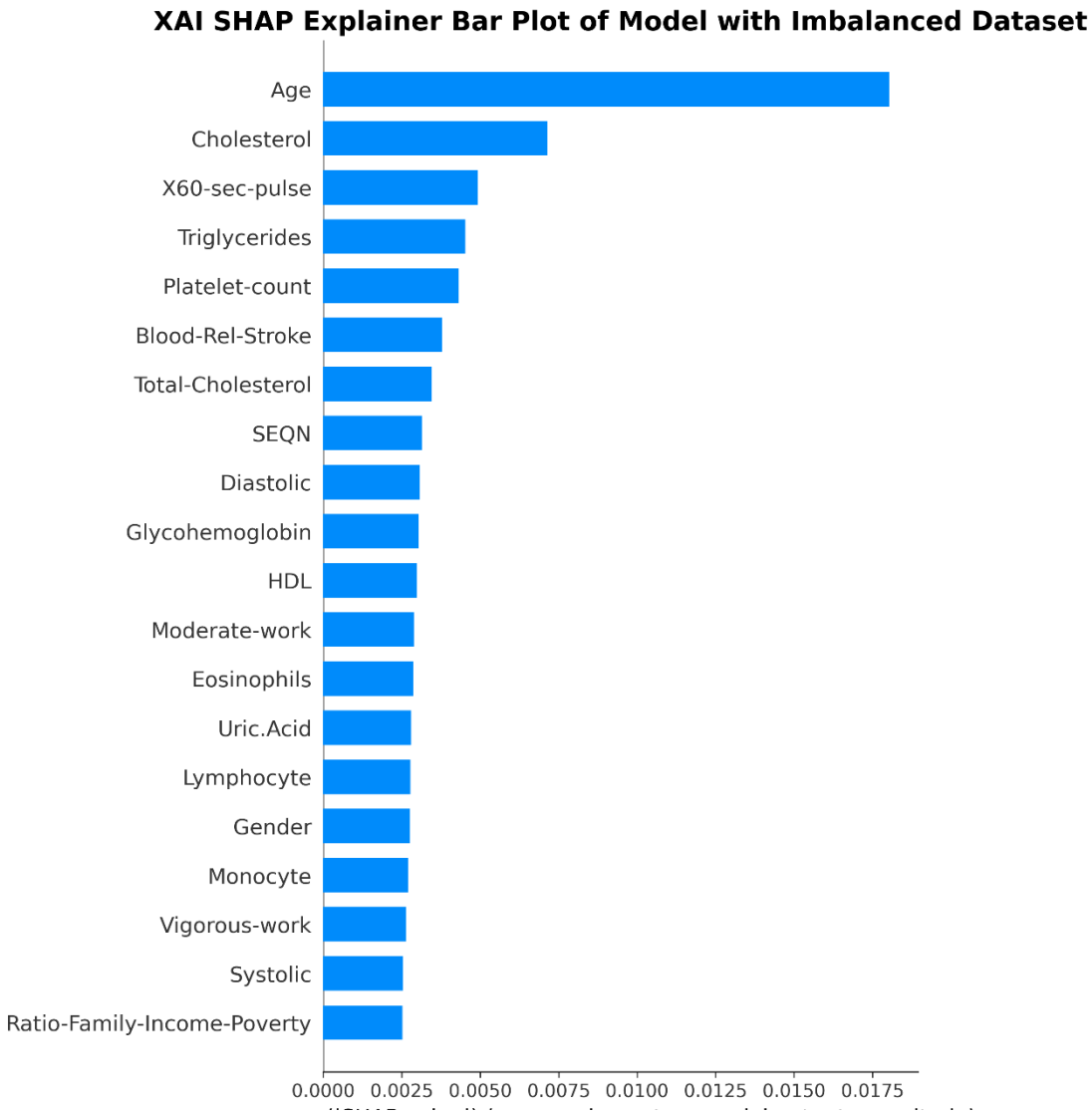

Figure 4.6: SHAP Feature Ranking Bar Plot (Imbalanced Dataset)

This model-agnostic analysis goes beyond traditional AI model evaluation measures. It offers insights that can be beneficial for refining the model computation time and other characteristics, addressing class imbalances in datasets, and diagnosing the performance of specific classes. The use of SHAP analysis along with standard evaluation metrics offers a complete way to understand, interpret, and improve model performance in complex, imbalanced situations.

4.3.2 SHAP Plots (Balanced Dataset)

By applying XAI, the SHAP algorithm helps to recognize the model decision process without needing changes to the model itself. Through SHAP values, the study expressed a ranked list of features based on their impact on decision-making process see in Table 4.4. The top features include Age, Blood Relative Stroke, and Diabetes, followed by Cholesterol, Moderate Work, and others see Figure 4.7 and Figure 4.8. These rankings give a clear view of which features influence predictions the most.

Table 4.4: SHAP values of top 20 features (Balanced Dataset)

S/N	Feature Name	SHAP Value (average)
1	Age	0.121196070

2	Blood-Rel-Stroke	0.066004942
3	Diabetes	0.052924895
4	Cholesterol	0.052186109
5	Moderate-work	0.046626729
6	Blood-Rel-Diabetes	0.041613206
7	Triglycerides	0.034731608
8	Albumin	0.032534138
9	Height	0.032253639
10	Platelet-count	0.030990272
11	X60-sec-pulse	0.029875804
12	Bilirubin	0.026597901
13	LDH	0.025012983
14	Monocyte	0.023612236
15	Hematocrit	0.023432326
16	Creatinine	0.023343070
17	Protein	0.022641045
18	Basophils	0.021600785
19	White-Blood-Cells	0.019542482
20	Uric.Acid	0.019265636

In the SHAP summary plot, the red color presents the higher value of the feature and the blue color presents the low value of the feature in the column. Through this visualization, the researchers see the contribution and effect of feature value on the overall feature ranking according to the architecture weight matrix.

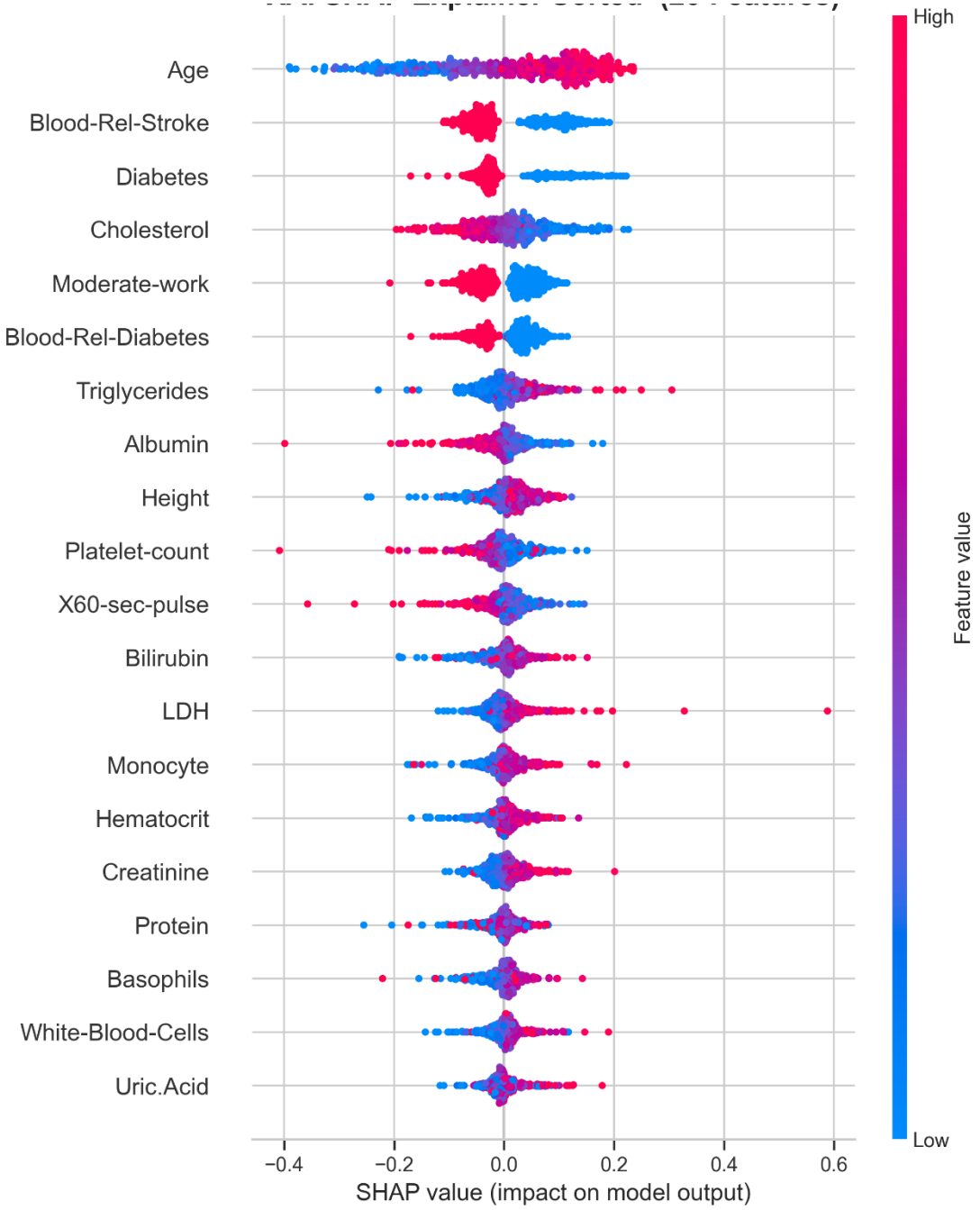

Figure 4.7: SHAP Feature Ranking Summary Plot (Balanced Dataset)

These findings go beyond simply making the black boxes understandable, but they have real value in the medical field. SHAP values add reliability to the AI model by visibly displaying the factors that affect its prediction process. This clearness can help medical professionals and researchers recognize what feature values drive each prediction class. These insights potentially guide clinical decisions or suggesting about the

lifestyle changes of patients. Future work could involve improving the model based on these key features or exploring more data sources to gain even deeper insights

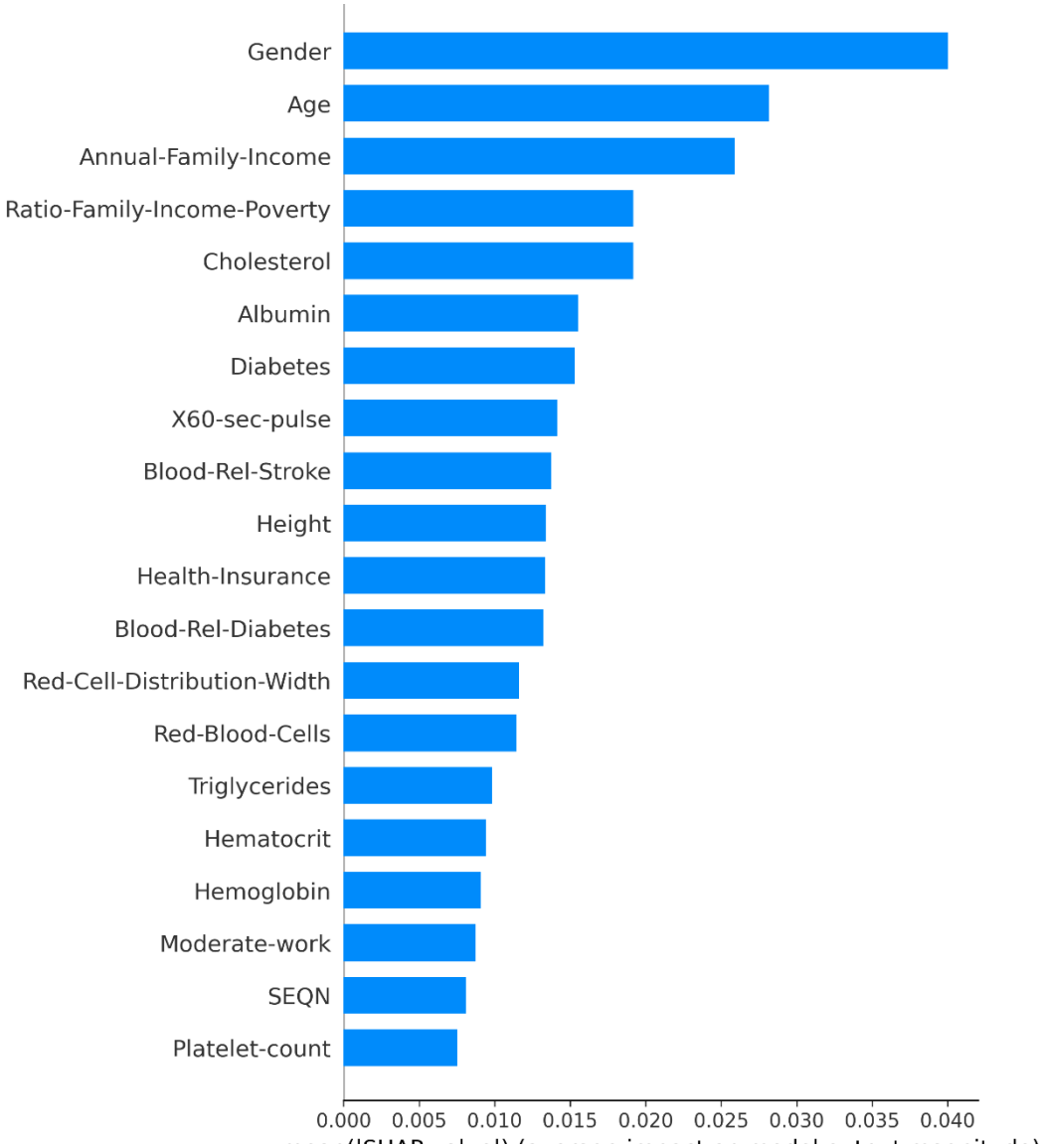

Figure 4.8: SHAP Feature Ranking Bar Plot (Balanced Dataset)

The application of the SHAP algorithm in this research study not only increases model transparency and trustworthiness but also creates opportunities for ongoing enhancement and a deeper knowledge of complex feature-prediction relationships in the medical field.

Chapter 5: Conclusion and Future Work

In conclusion, this empirical study has attempted to explore the area of cardiovascular health classification with the help of the Multi-Layer Perceptron neural network model. The proposed model achieves admirable accuracy, recall, and precision. Improving predictive models for heart health necessitates ongoing research into diverse types of neural networks. This means exploring various designs and structures to make the models more accurate, precise, and implementable.

The next step in the research should explore the combining of different CNNs, RNNs, LSTMs, and the BiLSTMs networks. CNNs are great at finding patterns in spatial data, which makes them useful to explore medical datasets. RNN networks process data in sequence, making them useful for understanding time-related health records of patients. LSTMs are the next generation of RNNs. LSTM networks solve the vanishing gradient problem and remember sequential information better. This capability of LSTMs makes them ideal for long-term health monitoring and forecasting the problems timely.

Future research will focus on exploring different medical datasets to increase the model's range and expand its capability to perform more accurately in classification and forecasting with various types of data. Datasets like Statlog, Cleveland, Hungarian Heart Disease, and Framingham [82] offer exceptional characteristics and challenges, allowing a more detailed evaluation of the model in robustness across various patient cohorts. These datasets take forth variations in demographic profiles, risk factors, and healthcare observations, and provide a chance to confirm the flexibility of the model in real-world situations.

There is also a gap in enhancing the model interpretability through other XAI techniques like DeepSHAP [94], DeepLIFT [95], Local Interpretable Model-agnostic Explanations (LIME) [96], FairML, and Causal Explanations (CXplain) [97]. These techniques play the fundamental role of evaluating the biases of the model and transparency about the decision-making process. Applying these XAI algorithms can give us more useful information like top feature contributions about the particular data instance with respect to predicted class. CXplain provides an understanding of decision boundaries about the model, enabling a nuanced understanding of intricate data patterns. The model-agnostic technique of XAI which is called LIME, makes understanding models easier by tweaking instances and examining responses or predicted classes. FairML technique makes sure that models are ethical and unbiased, which is essential in the healthcare system. At the end, the study concludes that these XAI methods will involve careful evaluation of the model's decision-making, key features, and ethical impact. The transparency they offer meets healthcare standards, helping build trust in AI-based systems.

References

[1] B. S and R. P, “Impact of Deep Learning Algorithms in Cardiovascular Disease Prediction,” *NVEO - Nat. VOLATILES Essent. OILS J. NVEO*, pp. 4341–4353, Nov. 2021.

[2] Y. Li *et al.*, “Geochemical Characteristics and Significance of Organic Matter in Hydrate-Bearing Sediments from Shenhu Area, South China Sea,” *Molecules*, vol. 27, no. 8, p. 2533, Apr. 2022, doi: 10.3390/molecules27082533.

[3] N. Zemzemi, S. Labarthe, R. D. Dubois, and Y. Coudière, “From body surface potential to activation maps on the atria: A machine learning technique,” in *2012 Computing in Cardiology*, Sep. 2012, pp. 125–128.

[4] A. Malik, T. Peng, and M. L. Trew, “A machine learning approach to reconstruction of heart surface potentials from body surface potentials,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI: IEEE, Jul. 2018, pp. 4828–4831. doi: 10.1109/EMBC.2018.8513207.

[5] J. Horvath, L. Shien, T. Peng, A. Malik, M. Trew, and L. Bear, “Deep learning neural nets for detecting heart activity,” Feb. 05, 2019, *arXiv*: arXiv:1901.09831. Accessed: Nov. 10, 2023. [Online]. Available: http://arxiv.org/abs/1901.09831

[6] L. R. Bear, P. R. Huntjens, R. D. Walton, O. Bernus, R. Coronel, and R. Dubois, “Cardiac electrical dyssynchrony is accurately detected by noninvasive electrocardiographic imaging,” *Heart Rhythm*, vol. 15, no. 7, pp. 1058–1069, Jul. 2018, doi: 10.1016/j.hrthm.2018.02.024.

[7] G. Hinton, S. Osindero, M. Welling, and Y.-W. Teh, “Unsupervised Discovery of Nonlinear Structure Using Contrastive Backpropagation,” *Cogn. Sci.*, vol. 30, no. 4, pp. 725–731, Jul. 2006, doi: 10.1207/s15516709cog0000_76.

[8] R. K. Bhagat, A. Yadav, Y. K. Rajoria, S. Raj, and R. Boadh, “Study of Fuzzy and Artificial Neural Network (ANN) Based Techniques to Diagnose Heart Disease,” *J. Pharm. Negat. Results*, vol. 13, no. 5, 2022.

[9] H. Ide and T. Kurita, “Improvement of learning for CNN with ReLU activation by sparse regularization,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA: IEEE, May 2017, pp. 2684–2691. doi: 10.1109/IJCNN.2017.7966185.

[10] T. Kobayashi, “Global Feature Guided Local Pooling,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 3364–3373. doi: 10.1109/ICCV.2019.00346.

[11] D. Herndon, F. Zhang, and W. Lineaweaver, “Metabolic Responses to Severe Burn Injury,” *Ann. Plast. Surg.*, vol. 88, no. 2, p. S128, Apr. 2022, doi: 10.1097/SAP.0000000000003142.

[12] A. E. Stoica, C. Chircov, and A. M. Grumezescu, “Hydrogel Dressings for the Treatment of Burn Wounds: An Up-To-Date Overview,” *Materials*, vol. 13, no. 12, p. 2853, Jun. 2020, doi: 10.3390/ma13122853.

[13] C. Crouzet, J. Q. Nguyen, A. Ponticorvo, N. P. Bernal, A. J. Durkin, and B. Choi, “Acute discrimination between superficial-partial and deep-partial thickness burns in a preclinical model with laser speckle imaging,” *Burns*, vol. 41, no. 5, pp. 1058–1063, Aug. 2015, doi: 10.1016/j.burns.2014.11.018.

[14] S. A. Suha and T. F. Sanam, “A deep convolutional neural network-based approach for detecting burn severity from skin burn images,” *Mach. Learn. Appl.*, vol. 9, p. 100371, Sep. 2022, doi: 10.1016/j.mlwa.2022.100371.

[15] Z. Ren *et al.*, “Deep attention-based neural networks for explainable heart sound classification,” *Mach. Learn. Appl.*, vol. 9, p. 100322, Sep. 2022, doi: 10.1016/j.mlwa.2022.100322.

[16] Y. Yu, X. Si, C. Hu, and J. Zhang, “A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures,” *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi: 10.1162/neco_a_01199.

[17] Ö. B. Mercan, S. N. Cavsak, A. Deliahmetoglu, and S. Tanberk, “Abstractive Text Summarization for Resumes With Cutting Edge NLP Transformers and LSTM,” in *2023 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Sivas, Turkiye: IEEE, Oct. 2023, pp. 1–6. doi: 10.1109/ASYU58738.2023.10296563.

[18] Y. Lin, Z. Chen, and Y. Yang, “Dynamic Forest Management Plan Selection and Optimization Based on Improved NLP, LSTM, and XGBoost,” In Review, preprint, Apr. 2023. doi: 10.21203/rs.3.rs-2770201/v1.

[19] J. Jo, J. Kung, and Y. Lee, “Approximate LSTM Computing for Energy-Efficient Speech Recognition,” *Electronics*, vol. 9, no. 12, p. 2004, Nov. 2020, doi: 10.3390/electronics9122004.
[20] Y. Li *et al.*, “BEHRT: Transformer for Electronic Health Records,” *Sci. Rep.*, vol. 10, no. 1, p. 7155, Apr. 2020, doi: 10.1038/s41598-020-62922-y.
[21] L. Wang, “Deep Learning Techniques to Diagnose Lung Cancer,” *Cancers*, vol. 14, no. 22, p. 5569, Nov. 2022, doi: 10.3390/cancers14225569.
[22] T. Lluka and J. M. Stokes, “Antibiotic discovery in the artificial intelligence era,” *Ann. N. Y. Acad. Sci.*, vol. 1519, no. 1, pp. 74–93, Jan. 2023, doi: 10.1111/nyas.14930.
[23] K. Preuss *et al.*, “Using Quantitative Imaging for Personalized Medicine in Pancreatic Cancer: A Review of Radiomics and Deep Learning Applications,” *Cancers*, vol. 14, no. 7, p. 1654, Mar. 2022, doi: 10.3390/cancers14071654.
[24] A. A. Nancy, D. Ravindran, P. M. D. Raj Vincent, K. Srinivasan, and D. Gutierrez Reina, “IoT-Cloud-Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning,” *Electronics*, vol. 11, no. 15, p. 2292, Jul. 2022, doi: 10.3390/electronics11152292.
[25] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, “Explainable AI: A Review of Machine Learning Interpretability Methods,” *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/e23010018.
[26] L. Weber, S. Lapuschkin, A. Binder, and W. Samek, “Beyond explaining: Opportunities and challenges of XAI-based model improvement,” *Inf. Fusion*, vol. 92, pp. 154–176, Apr. 2023, doi: 10.1016/j.inffus.2022.11.013.
[27] A. Adadi and M. Berrada, “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),” *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.
[28] S. S Band *et al.*, “Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods,” *Inform. Med. Unlocked*, vol. 40, p. 101286, Jan. 2023, doi: 10.1016/j.imu.2023.101286.
[29] C. Manresa-Yee, M. F. Roig-Maimó, S. Ramis, and R. Mas-Sansó, “Advances in XAI: Explanation Interfaces in Healthcare,” in *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects*, C.-P. Lim, Y.-W. Chen, A. Vaidya, C. Mahorkar, and L. C. Jain, Eds., in Intelligent Systems Reference Library. , Cham: Springer International Publishing, 2022, pp. 357–369. doi: 10.1007/978-3-030-83620-7_15.
[30] X. Dai, M. T. Keane, L. Shalloo, E. Ruelle, and R. M. J. Byrne, “Counterfactual Explanations for Prediction and Diagnosis in XAI,” in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, in AIES ’22. New York, NY, USA: Association for Computing Machinery, Jul. 2022, pp. 215–226. doi: 10.1145/3514094.3534144.
[31] S. Mertes, T. Huber, K. Weitz, A. Heimerl, and E. André, “GANterfactual—Counterfactual Explanations for Medical Non-experts Using Generative Adversarial Learning,” *Front. Artif. Intell.*, vol. 5, 2022, doi: 10.3389/frai.2022.825565.
[32] J. Duell, X. Fan, B. Burnett, G. Aarts, and S.-M. Zhou, “A Comparison of Explanations Given by Explainable Artificial Intelligence Methods on Analysing Electronic Health Records,” in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Jul. 2021, pp. 1–4. doi: 10.1109/BHI50953.2021.9508618.
[33] D. W. Joyce, A. Kormilitzin, K. A. Smith, and A. Cipriani, “Explainable artificial intelligence for mental health through transparency and interpretability for understandability,” *Npj Digit. Med.*, vol. 6, no. 1, Art. no. 1, Jan. 2023, doi: 10.1038/s41746-023-00751-9.
[34] I. Vaccari, V. Orani, A. Paglialonga, E. Cambiaso, and M. Mongelli, “A Generative Adversarial Network (GAN) Technique for Internet of Medical Things Data,” *Sensors*, vol. 21, no. 11, Art. no. 11, Jan. 2021, doi: 10.3390/s21113726.
[35] H. Ahmed, E. M. G. Younis, A. Hendawi, and A. A. Ali, “Heart disease identification from patients’ social posts, machine learning solution on Spark,” *Future Gener. Comput. Syst.*, vol. 111, pp. 714–722, Oct. 2020, doi: 10.1016/j.future.2019.09.056.
[36] J. Rashid, S. Kanwal, J. Kim, M. Wasif Nisar, U. Naseem, and A. Hussain, “Heart Disease Diagnosis Using the Brute Force Algorithm and Machine Learning Techniques,” *Comput. Mater. Contin.*, vol. 72, no. 2, pp. 3195–3211, 2022, doi: 10.32604/cmc.2022.026064.

[37] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.
[38] J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction," *Int. J. Comput. Appl.*, vol. 17, no. 8, pp. 43–48, Mar. 2011, doi: 10.5120/2237-2860.
[39] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Comput. Biol. Med.*, vol. 136, p. 104672, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104672.
[40] D. Shah, S. Patel, and S. K. Bharti, "Heart Disease Prediction using Machine Learning Techniques," *SN Comput. Sci.*, vol. 1, no. 6, p. 345, Oct. 2020, doi: 10.1007/s42979-020-00365-y.
[41] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," *Algorithms*, vol. 16, no. 2, p. 88, Feb. 2023, doi: 10.3390/a16020088.
[42] A. Pandita, "Prediction of Heart Disease using Machine Learning Algorithms," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 9, no. VI, pp. 2422–2429, Jun. 2021, doi: 10.22214/ijraset.2021.3412.
[43] A. Lakshmanarao, Y. Swathi, and P. S. S. Sundareswar, "Machine learning techniques for heart disease prediction," *Forest*, vol. 95, no. 99, p. 97, 2019.
[44] L. Yahaya, N. David Oye, and E. Joshua Garba, "A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques," *Am. J. Artif. Intell.*, vol. 4, no. 1, p. 20, 2020, doi: 10.11648/j.ajai.20200401.12.
[45] F. S. Alotaibi, "Implementation of Machine Learning Model to Predict Heart Failure Disease," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 10, no. 6, Art. no. 6, 29 2019, doi: 10.14569/IJACSA.2019.0100637.
[46] N. Bora, S. Gutta, and A. Hadaegh, "Using Machine Learning to Predict Heart Disease (Review Paper)," *WSEAS Trans. Biol. Biomed.*, vol. 19, pp. 1–9, Jan. 2022, doi: 10.37394/23208.2022.19.1.
[47] H. Ayatollahi, L. Gholamhosseini, and M. Salehi, "Predicting coronary artery disease: a comparison between two data mining algorithms," *BMC Public Health*, vol. 19, no. 1, p. 448, Apr. 2019, doi: 10.1186/s12889-019-6721-5.
[48] B. U. Rindhe, N. Ahire, R. Patil, S. Gagare, and M. Darade, "Heart disease prediction using machine learning," *Heart Dis.*, vol. 5, no. 1, 2021.
[49] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," *Inform. Med. Unlocked*, vol. 26, p. 100655, 2021, doi: 10.1016/j.imu.2021.100655.
[50] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, "Heart Diseases Diagnosis Using Neural Networks Arbitration," *Int. J. Intell. Syst. Appl.*, vol. 7, no. 12, p. 75, doi: 10.5815/ijisa.2015.12.08.
[51] P. Ghadge, V. Girme, K. Kokane, and P. Deshmukh, "Intelligent Heart Attack Prediction System Using Big Data," vol. 2, no. 2, 2015.
[52] A. Rajkumar and M. G. S. Reena, "Diagnosis Of Heart Disease Using Datamining Algorithm," 2010.
[53] A. Hazra, S. K. Mandal, A. Gupta, A. Mukherjee, and A. Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review," 2017.
[54] J. Jonnagaddala, S.-T. Liaw, P. Ray, M. Kumar, N.-W. Chang, and H.-J. Dai, "Coronary artery disease risk assessment from unstructured electronic health records using text mining," *J. Biomed. Inform.*, vol. 58, pp. S203–S210, Dec. 2015, doi: 10.1016/j.jbi.2015.08.003.
[55] N. Hasan and Y. Bao, "Comparing different feature selection algorithms for cardiovascular disease prediction," *Health Technol.*, vol. 11, no. 1, pp. 49–62, Jan. 2021, doi: 10.1007/s12553-020-00499-2.
[56] C. S.Dangare and S. S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques," *Int. J. Comput. Appl.*, vol. 47, no. 10, pp. 44–48, Jun. 2012, doi: 10.5120/7228-0076.
[57] J. Nahar, T. Imam, K. S. Tickle, and Y.-P. P. Chen, "Association rule mining to detect factors which contribute to heart disease in males and females," *Expert Syst. Appl.*, vol. 40, no. 4, pp. 1086–1093, Mar. 2013, doi: 10.1016/j.eswa.2012.08.028.
[58] L. Baccour, "Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets," *Expert Syst. Appl.*, vol. 99, pp. 115–125, Jun. 2018, doi: 10.1016/j.eswa.2018.01.025.

[59] M. Shamsollahi, A. Badiee, and M. Ghazanfari, “Using Combined Descriptive and Predictive Methods of Data Mining for Coronary Artery Disease Prediction: a Case Study Approach,” *J. AI Data Min.*, vol. 7, no. 1, pp. 47–58, Jan. 2019, doi: 10.22044/jadm.2017.4992.1599.
[60] Z. Al-Makhadmeh and A. Tolba, “Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: A classification approach,” *Measurement*, vol. 147, p. 106815, Dec. 2019, doi: 10.1016/j.measurement.2019.07.043.
[61] A. Akella and S. Akella, “Machine learning algorithms for predicting coronary artery disease: efforts toward an open source solution,” *Future Sci. OA*, vol. 7, no. 6, p. FSO698, Jul. 2021, doi: 10.2144/fsoa-2020-0206.
[62] R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7675–7680, May 2009, doi: 10.1016/j.eswa.2008.09.013.
[63] C.-A. Cheng and H.-W. Chiu, “An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo: IEEE, Jul. 2017, pp. 2566–2569. doi: 10.1109/EMBC.2017.8037381.
[64] C.-Y. Hsieh *et al.*, “Taiwan’s National Health Insurance Research Database: past and future,” *Clin. Epidemiol.*, vol. 11, pp. 349–358, May 2019, doi: 10.2147/CLEP.S196293.
[65] S. Zaman and R. Toufiq, “Codon based back propagation neural network approach to classify hypertension gene sequences,” in *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Cox’s Bazar, Bangladesh: IEEE, Feb. 2017, pp. 443–446. doi: 10.1109/ECACE.2017.7912945.
[66] K. Subhadra and B. Vikas, “Neural network based intelligent system for predicting heart disease,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 5, pp. 484–487, 2019.
[67] H. Meshref, “Cardiovascular Disease Diagnosis: A Machine Learning Interpretation Approach,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 12, 2019, doi: 10.14569/IJACSA.2019.0101236.
[68] T. F. Romdhane, H. Alhichri, R. Ouni, and M. Atri, “Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss,” *Comput. Biol. Med.*, vol. 123, p. 103866, Aug. 2020, doi: 10.1016/j.compbiomed.2020.103866.
[69] A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, “An efficient convolutional neural network for coronary heart disease prediction,” *Expert Syst. Appl.*, vol. 159, p. 113408, Nov. 2020, doi: 10.1016/j.eswa.2020.113408.
[70] Z. Du *et al.*, “Accurate Prediction of Coronary Heart Disease for Patients With Hypertension From Electronic Health Records With Big Data and Machine-Learning Methods: Model Development and Performance Evaluation,” *JMIR Med. Inform.*, vol. 8, no. 7, p. e17257, Jul. 2020, doi: 10.2196/17257.
[71] J. K. Kim and S. Kang, “Neural Network-Based Coronary Heart Disease Risk Prediction Using Feature Correlation Analysis,” *J. Healthc. Eng.*, vol. 2017, p. e2780501, Sep. 2017, doi: 10.1155/2017/2780501.
[72] H. Yang, Z. Chen, H. Yang, and M. Tian, “Predicting Coronary Heart Disease Using an Improved LightGBM Model: Performance Analysis and Comparison,” *IEEE Access*, vol. 11, pp. 23366–23380, 2023, doi: 10.1109/ACCESS.2023.3253885.
[73] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, “Heart disease prediction using machine learning algorithms,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, p. 012072, Jan. 2021, doi: 10.1088/1757-899X/1022/1/012072.
[74] F. Ali *et al.*, “A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion,” *Inf. Fusion*, vol. 63, pp. 208–222, Nov. 2020, doi: 10.1016/j.inffus.2020.06.008.
[75] M. A. Khan, “An IoT Framework for Heart Disease Prediction Based on MDCNN Classifier,” *IEEE Access*, vol. 8, pp. 34717–34727, 2020, doi: 10.1109/ACCESS.2020.2974687.
[76] F. Jabeen *et al.*, “An IoT based efficient hybrid recommender system for cardiovascular disease,” *Peer--Peer Netw. Appl.*, vol. 12, no. 5, pp. 1263–1276, Sep. 2019, doi: 10.1007/s12083-019-00733-3.
[77] T. Menzies, E. Kocagüneli, L. Minku, F. Peters, and B. Turhan, “Chapter 24 - Using Goals in Model-Based Reasoning,” in *Sharing Data and Models in Software Engineering*, T. Menzies, E. Kocagüneli, L. Minku, F. Peters, and B. Turhan, Eds., Boston: Morgan Kaufmann, 2015, pp. 321–353. doi: 10.1016/B978-0-12-417295-1.00024-2.

[78] S. Tuli *et al.*, "HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments," *Future Gener. Comput. Syst.*, vol. 104, pp. 187–200, Mar. 2020, doi: 10.1016/j.future.2019.10.043.
[79] J. Kwon, K.-H. Kim, K.-H. Jeon, and J. Park, "Deep learning for predicting in-hospital mortality among heart disease patients based on echocardiography," *Echocardiography*, vol. 36, no. 2, pp. 213–218, 2019, doi: 10.1111/echo.14220.
[80] "Statlog (Heart) Data Set." [Online]. Available: https://www.kaggle.com/datasets/shubamsumbria/statlog-heart-data-set
[81] "Heart Disease Cleveland UCI." [Online]. Available: https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci
[82] "Framingham_CHD_preprocessed_data." [Online]. Available: https://www.kaggle.com/datasets/captainozlem/framingham-chd-preprocessed-data
[83] "NHANES Datasets." [Online]. Available: https://www.kaggle.com/datasets/mubashiriqbal07/cardiacdiseasepredictiondataset40x37079
[84] "Cardiovascular Disease dataset 70000." [Online]. Available: https://www.kaggle.com/sulianova/competitions
[85] A. M. Antoniadi *et al.*, "Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review," *Appl. Sci.*, vol. 11, no. 11, p. 5088, May 2021, doi: 10.3390/app11115088.
[86] S. Das, M. Sultana, S. Bhattacharya, D. Sengupta, and D. De, "XAI–reduct: accuracy preservation despite dimensionality reduction for heart disease classification using explainable AI," *J. Supercomput.*, vol. 79, no. 16, pp. 18167–18197, Nov. 2023, doi: 10/gss83j.
[87] V. Belle and I. Papantonis, "Principles and Practice of Explainable Machine Learning," *Front. Big Data*, vol. 4, p. 688969, Jul. 2021, doi: 10/gnjm43.
[88] L. Luotsinen, D. Oskarsson, P. Svenmarck, and U. W. Bolin, "Explainable Artificial Intelligence: Exploring XAI Techniques in Military Deep Learning Applications," 2019.
[89] D. Gunning and D. W. Aha, "DARPA's Explainable Artificial Intelligence Program," *AI Mag.*, vol. 40, no. 2, pp. 44–58, Jun. 2019, doi: 10/gh24wc.
[90] S. Laato, M. Tiainen, A. K. M. Najmul Islam, and M. Mäntymäki, "How to explain AI systems to end users: a systematic literature review and research agenda," *Internet Res.*, vol. 32, no. 7, pp. 1–31, Dec. 2022, doi: 10.1108/intr-08-2021-0600.
[91] P. Guleria, P. Naga Srinivasu, S. Ahmed, N. Almusallam, and F. K. Alarfaj, "XAI Framework for Cardiovascular Disease Prediction Using Classification Techniques," *Electronics*, vol. 11, no. 24, p. 4086, Dec. 2022, doi: 10/gss83p.
[92] M. Ahsan, "Heart Attack Prediction using Machine Learning and XAI," 2023.
[93] A. Nascita, A. Montieri, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapé, "Improving Performance, Reliability, and Feasibility in Multimodal Multitask Traffic Classification with XAI," *IEEE Trans. Netw. Serv. Manag.*, vol. 20, no. 2, pp. 1267–1289, Jun. 2023, doi: 10.1109/TNSM.2023.3246794.
[94] A. T. Keleko, B. Kamsu-Foguem, R. H. Ngouna, and A. Tongne, "Health condition monitoring of a complex hydraulic system using Deep Neural Network and DeepSHAP explainable XAI," *Adv. Eng. Softw.*, vol. 175, p. 103339, Jan. 2023, doi: 10.1016/j.advengsoft.2022.103339.
[95] S. Nazir, D. M. Dickson, and M. U. Akram, "Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks," *Comput. Biol. Med.*, vol. 156, p. 106668, Apr. 2023, doi: 10.1016/j.compbiomed.2023.106668.
[96] Y. Wu, L. Zhang, U. A. Bhatti, and M. Huang, "Interpretable Machine Learning for Personalized Medical Recommendations: A LIME-Based Approach," *Diagnostics*, vol. 13, no. 16, p. 2681, Aug. 2023, doi: 10.3390/diagnostics13162681.
[97] P. Schwab and W. Karlen, "CXPlain: Causal Explanations for Model Interpretation under Uncertainty," Oct. 2019.