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. This study undertakes the ambitious task of advancing predictive modeling for coronary heart disease through the innovative integration of attention-layer-based multi-layer perceptron (MLP) architecture network model. The dataset under scrutiny encompasses a rich tapestry of 51 distinct characteristics distributed across 37079 records, offering a comprehensive and nuanced depiction of diverse patient profiles. The heart of this study lies in the proposed model's architecture, where attention mechanisms play a pivotal role in capturing intricate patterns within the data. The MLP architecture, augmented with attention layers, exhibits a unique ability to focus on salient features and relationships, enhancing the model's capacity to discern complex patterns associated with coronary heart disease. The research journey commences with meticulous preprocessing procedures, transcending the mere compilation of data. Addressing missing information, eliminating duplicates, and deploying the Synthetic Minority Over-sampling Technique (SMOTE) collectively form an integral part of the data preparation phase. The inclusion of SMOTE not only rectifies inherent class imbalances but also enhances the model's robustness by oversampling minority instances. Following the architectural exposition, the network undergoes extensive training and meticulous evaluation, culminating in impressive outcomes. The model achieves a test accuracy of 97.10%, underscoring its efficacy in discerning instances of the target class. Sensitivity and precision, measuring 97.85% and 96.40% respectively, showcase the model's prowess in accurately identifying positive cases while minimizing false positives. Furthermore, the Area Under the Curve (AUC) metric serves as a testament to the model's discriminative power, further validating its clinical relevance. Beyond the numerical achievements, the interpretability of the model is a noteworthy feature, adding a layer of transparency to the decision-making process.

This research significantly contributes to the burgeoning field of cardiovascular health prediction, presenting a model that excels not only in accuracy but also in interpretability. The amalgamation of advanced attention mechanisms, diverse neural network architectures, and eXplainable Artificial Intelligence (XAI) techniques positions this study at the forefront of predictive modeling for critical healthcare applications. As the model proves its mettle in discerning intricate patterns within cardiovascular datasets, the implications for improved patient care and proactive intervention become increasingly tangible.

Table of Content

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 **Error! Reference source not found.**, 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 within a doughnut-shaped scanner. To obtain images of the heart and chest, the device rotates an X-ray tube around the human body.
- **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 prevalence of these risk factors is likely contributing to the high burden of CVD in Pakistan. The Pakistani government has recognized the importance of addressing CVD and has developed some initiatives to reduce the burden of disease. These initiatives include raising awareness of CVD and its risk factors, promoting healthy lifestyles, improving access to early detection and treatment services. However, more needs to be done to reduce the burden of CVD in Pakistan. This includes investing in research to develop better ways to prevent, diagnose, and treat CVD.

1.1 Background

Cardiac catheterization is a minimally invasive technique that utilizes a slender, pliable tube (catheter) to assess the heart's electrical activity. The catheter is inserted into a vein in the leg and navigated up to the heart, where it is positioned in the ventricles or atria. Electrodes at the catheter's tip measure the electrical signals from the cardiac muscle. Electrograms (EGMs) are the name given to these measurements. The Heart Surface Potential (HSP) signal at the specific location where the EGM was recorded is represented by the signal. Regretfully, there is a considerable chance of problems with these cardiac treatments. EGMs yield valuable information about the heart's rhythm and function. Numerous cardiac disorders, such as arrhythmias, heart failure, and coronary artery disease, can be identified with them. EGMs can also be used to direct specific cardiac treatments, such as ablation therapy for arrhythmias and angioplasty for coronary artery disease [3]. They may result in fatalities, heart attacks, and strokes [4]. Little toddlers have small cardiac chambers and narrow veins, making this EGM treatment more risky when done on them [5]. Cardiac electrical impulses propagate throughout the body via specialized tissues called cardiac muscle fibers [6]. When these impulses reach the skin's surface, they can be measured using electrodes. This serves as the basis for electrocardiography (ECG), a non-invasive method of capturing 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. A twelve-lead ECG places twelve electrodes on various body parts to capture the heart's electrical activity from twelve distinct angles. This gives a clearer picture of the electrical activity of the heart than a single-lead ECG. Patients undergoing cardiac resynchronization therapy (CRT), a procedure that uses a pacemaker to synchronize the heart's ventricles, are often chosen based on 12-lead ECG data. CRT is particularly beneficial for patients with heart failure, a condition in which the heart muscle is weak and unable to pump blood as efficiently as it could. However, approximately one-third of CRT patients do not have a favorable response to treatment. This is most likely because ECG data is not always able to precisely locate and measure electrical inhomogeneities in the

heart. Electrical inhomogeneities are areas of the heart where the electrical activity is abnormal. The heart's capacity to pump blood efficiently may be hampered by certain anomalies.

Artificial neural networks (ANNs) are used in deep learning (DL), a section of machine learning (ML), to learn from labeled data. ANNs can recognize complicated patterns from data without explicit programming, and they are inspired by the structure and operations of the human brain. In a variety of applications, such as Machine Translation (MT), Natural Language Processing (NLP), and Image Recognition (IR), DL has produced cutting-edge outcomes. Additionally, it is being utilized to create fresh and inventive applications in industries like robotics, healthcare, finance, and many more. DL originated in the 1940s when scientists started working on ANNs. However, it wasn't until the early 2000s that DL started to receive a lot of attention, largely because of developments in learning algorithms and processing power. In 2006, Geoffrey Hinton published a research article [7] that showed how to train deep neural networks using a technique called backpropagation. This discovery opened the door for the creation of contemporary DL algorithms. The first-ever ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [8] was won by a DL model in 2012. This event marked a watershed moment in the development of DL, and it demonstrated the potential of DL to solve real-world problems. There are a wide variety of DL techniques 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 ANNs. ANNs are used in DL, a kind of ML, to extract knowledge from data. DL models don't usually need to be explicitly coded in order 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 and other applications using sequential data are a good fit for recurrent neural networks (RNNs). Time-series data can be used by RNNs to extract sequential representations through the use of a network of recurrent layers. The recurrent units that comprise each recurrent layer are in charge of processing the relevant time step of the incoming data. The hidden states of each recurrent layer are carried over into the subsequent recurrent layer. Lastly, the target classes are predicted using the hidden states of the last recurrent layer [15].

Long Short-Term Memory(LSTM) is a type of artificial RNN design used in the field of DL [16]. Text, audio, and time series data are among the types of sequences of data that LSTMs excel at processing and forecasting. NLP is among the many tasks to which they have been successfully used [17], speech recognition, and machine translation. The vanishing gradient problem, a frequent difficulty with RNNs that makes it challenging to identify long-term dependencies in data, is solved by this variation of the traditional RNN architecture. A series of recurrently connected LSTM cells make up an LSTM network. Every LSTM cell is made up of four major parts: The information from the previous cell state that should be forgotten is selected by the forget gate. The input gate is responsible for selecting which new data to add to the cell state. Cell state: This is the cell's memory, where it keeps track of the knowledge it has accumulated throughout time. The output gate selects the information that will be output from the cell state. The network can discover long-term dependencies in data since the LSTM architecture permits information to travel both forward and backward across the network. LSTM networks have shown to be effective at learning long-term dependencies from data. With them, cutting-edge outcomes on a variety of jobs have been attained, including NLP [18], speech recognition [19], and machine translation.

Recently, transformer networks, a kind of neural network, have produced state-of-the-art outcomes in a variety of applications, such as NLP, image recognition, and machine translation. In study [20], The authors employ a transformer network model for electronic health records (EHRs) that is intended to be scalable, interpretable, and tailored for a broad range of illnesses and EHR modalities. After pre-training on a sizable dataset, Bidirectional Encoder Representations from Transformers (BEHRT) can be adjusted for particular

downstream tasks. In order to show off this feature, the scientists trained and tested the model to predict the next most likely diseases in a patient's future visits on Clinical Practice Research Datalink (CPRD), one of the most linked primary care EHR systems. BEHRT is adaptable enough to include more EHR data modalities because to its modular architecture.

DL is being used to develop new algorithms for medical image analysis. Through the use of these algorithms, radiologists can more quickly and reliably diagnose conditions and abnormalities in medical pictures. For example, DL models have been used to develop algorithms that can detect lung cancer in medical images with greater accuracy than human radiologists [21]. DL is being used to accelerate the drug discovery process. DL models can be used to screen millions of potential drug candidates for their efficacy and safety. This can help researchers to identify new drugs and treatments more quickly and efficiently. For example, DL models have been used to develop new drugs that are effective against antibiotic-resistant bacteria [22]. Patients can have individualized treatment regimens created for them using DL. To forecast a patient's response to various therapies, DL models can analyze genetic information, medical history, and other variables. With this information, a personalized treatment plan based on the patient's unique needs can be developed. In particular, pancreatic cancer patients' individualized treatment regimens have been created using DL models [23]. DL can be used to develop remote patient monitoring systems. These systems can be used to monitor patients' vital signs, activity levels, and other health data remotely. This can help to identify early warning signs of health problems and allow patients to receive treatment more quickly. For example, DL models have been used to develop remote patient monitoring systems for patients with heart disease [24].

1.2 eXplainable Artificial Intelligence (XAI)

AI has advanced significantly, with algorithms capable of tackling complicated problems and producing spectacular outcomes in a variety of disciplines. Even with these developments, AI models sometimes lack interpretability and openness, which makes it challenging to understand how they make decisions and the underlying reasons of their outputs. This lack of explainability raises issues about AI systems' trustworthiness, accountability, and justice. The increasing complexity and widespread adoption of AI models have brought to light the critical need for explainable or interpretable artificial intelligence (XAI) [25]. The goal of XAI is to give users visibility into the inner workings of AI models so they can comprehend the reasoning behind the model’s predictions and how they make judgments. This transparency is crucial for building trust in AI systems, ensuring accountability, and preventing potential biases or unfair outcomes. The concept of XAI can be dated to the initial stages of AI research [26]. In the 1980s, researchers began exploring methods for explaining the decisions of expert systems, laying the foundation for XAI. These early efforts focused on rule-based expert systems, which were relatively simple and easier to understand compared to modern AI models. In the 1990s, the development of ML algorithms and the increasing complexity of AI models led to a renewed interest in XAI [27]. Researchers began exploring techniques for explaining the predictions of ML models, such as DTs, rule lists, and sensitivity analysis. These methods provided some level of transparency but were often limited in their ability to explain complex models. The field of XAI has advanced significantly in recent

years, driven by the increasing complexity of AI models and the growing demand for explainable solutions. Researchers have developed a wide range of XAI techniques, each with its own strengths and limitations. Some of the XAI approaches that have been developed are shown in Figure 1.1. These are also explained in some detail below.

1.2.1 Anchor Explanation

Anchor explanations are a type of XAI technique that provides local explanations for individual predictions [28]. They aim to identify a small set of features or data points that are most responsible for the model's prediction, acting as “anchors” that explain the decision.

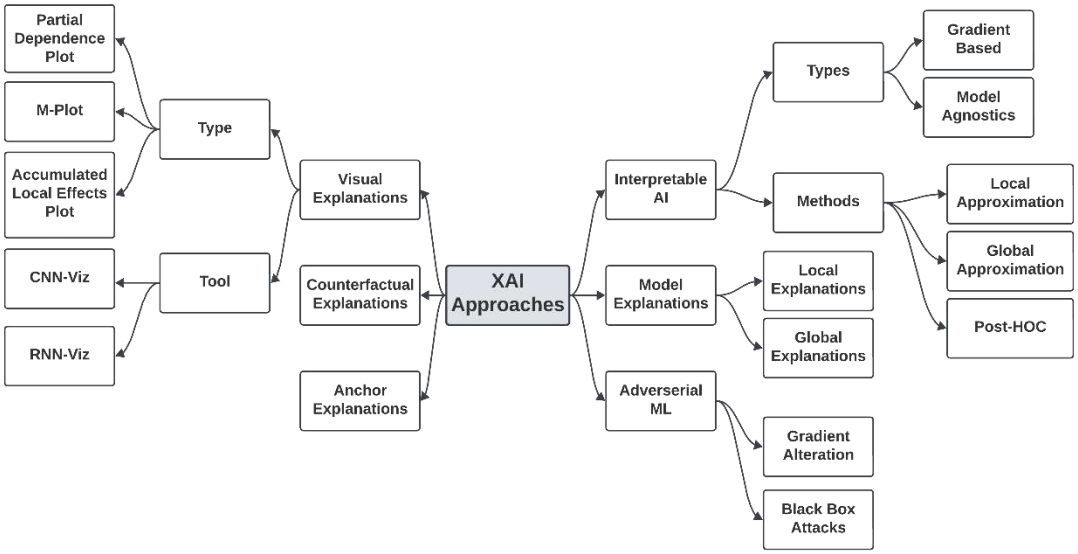

Figure 1.1: eXplainable AI Approaches

Anchor explanations are particularly useful for explaining complex models, such as deep neural networks, where it can be difficult to understand the contribution of individual features.

1.2.2 Visual Explanation

Visual explanations are a powerful tool for communicating XAI insights to users [29]. They use visual elements such as charts, graphs, and images to represent complex relationships between features and predictions, making them more intuitive and easier to understand than text-based explanations. Both local and global explanations can be better understood with the use of visual aids, which provide light on specific predictions as well as the general 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

Under the umbrella of XAI, interpretable AI (IAI) is the study of creating AI models with intrinsic interpretability [32], meaning that their decision-making processes can be understood without the need for external explanation techniques. This is a challenging goal, as it requires the development of new AI algorithms [33] that are designed to be transparent and interpretable from the outset.

1.2.5 Adversarial ML

Adversarial ML is a field that focuses on developing methods for attacking and manipulating AI models. Adversarial ML examples, or skillfully constructed inputs that deceive AI models into generating inaccurate predictions, can be produced through the application of adversarial machine learning algorithms. Adversarial ML is a powerful tool for testing the robustness and security of AI systems, and it can be utilized for develop new XAI techniques [34] by identifying how AI models can be deceived.

XAI can provide several benefits, including: *Improved understanding of AI models:* Researchers may make better judgments and comprehend how AI models operate with the aid of visualizations. *Increased trust in AI systems:* By providing transparency, visual explanations can help to build trust in AI systems. *Identification of potential biases:* Visualizations can help to identify potential biases in AI models by highlighting patterns in the data that may not be apparent from text-based explanations. *More informed decision-making:* Visual explanations can help users to make more informed decisions by providing insights into the factors that influence the model's predictions.

1.3 Problem Formulation and Problem Statement

Many AI models are being developed in the modern field of AI to aid in the medical industry. A lot of ML and DL models are being created to help physicians and other health care providers in disease prevention, prediction, and learning how to reduce the chances by prescribing 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 endeavors to grapple with the multifaceted challenges intrinsic to classification tasks when confronted with imbalanced datasets. By recognizing the class imbalances, and their tangible implications in real-world scenarios, the study advocates for a comprehensive approach. This involves data preprocessing techniques, model refinement strategies, and harnessing the interpretability provided by advanced methodologies like SHAP analysis. Through these concerted efforts, the goal is to bolster the resilience and dependability of DL models, particularly when navigating the complexities introduced by imbalanced data distributions.

1.4.1 Addressing Class Imbalance

Implementing strategies to address class imbalance is imperative. Techniques such as oversampling the minority class, employing advanced methods like SMOTE can enhance model performance by balancing the dataset.

1.4.2 Fine-Tuning Model Architecture

The model's capacity to identify patterns in both balanced and unbalanced data can be improved by iteratively fine-tuning the model architecture, experimenting with hyperparameter tuning, and investigating various options for the number of Dense Layers, number of nodes in different layers, number of training epochs, and batch size of data.

1.4.3 Utilizing SHAP Analysis

Applying SHAP algorithms allows for a granular understanding of feature importance in the model's decision-making process. Insights gained from SHAP analysis can guide feature engineering and further model refinement, particularly in addressing imbalances.

This research will boost trust in AI systems and demonstrate the transparency and understandability of AI systems in the health industry.

1.5 Research Contribution

In this research contribution, the study addresses challenges in classification arising from imbalanced datasets. The research objectives encompass a deep exploration of these challenges, innovative solution proposals, and a meaningful contribution to the existing knowledge landscape. The research carefully applies a set of comprehensive methodologies tailored to the complexities of imbalanced datasets. These methodologies guide the investigation into unexplored territories within the field, laying a solid foundation for the study. The findings encapsulate key insights, confirming existing knowledge and unveiling novel discoveries. The novelty lies in innovative methodologies applied to uncover patterns within imbalanced datasets, contributing significantly to the field. Theoretical contributions extend and challenge existing frameworks, providing a nuanced understanding of how ML and DL models interact with imbalanced data distributions. This research

fills gaps in understanding, complementing and diverging from current discourse. It has a significant impact on the direction of future research, enhancing theoretical understanding while having practical significance.

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 both 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, it was noted that the mainstream of studies focused on the Cleveland dataset, 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 dimentions, 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.

Yahaya et al [44] uses data mining and ML approaches to examine modern Clinical Decision Support Systems (CDSS) for the prediction of cardiac disease. Classification algorithms that are often used, such NB, DT, and ANN, have produced results with differing accuracy. The creation of predictive models for individuals with heart disease has yielded very limited results, underscoring the necessity for more complex models that use a variety of geographic data sources in order to increase the accuracy of early disease beginning prediction.

The research conducted by Alotaibi [45] involves a comprehensive investigation into subject disease classification and prediction utilizing ML techniques. The study employs algorithms such as NB, DT, RF, SVM, and LR within the RapidMiner framework. The analysis utilizes the widely adopted Cleveland heart disease dataset from the UCI ML repository, comprising 303 instances and 14 attributes. The model's learning and evaluation employ the 10-fold cross-validation method. Results indicate that the DT algorithm exhibits the highest accuracy in subject disease prediction, followed by SVM at 93.19% and 92.30%, respectively. The research presents a model combining five algorithms within the RapidMiner tool, demonstrating higher accuracy than Matlab and Weka. Despite the limitations of a small dataset, the study indicates significant improvements over previous research, emphasizing the potential utility for timely diagnoses in the medical field.

Bora et al [46] address the challenge through the proposition of heart disease prediction utilizing diverse ML algorithms, including LR, NB, SVM, KNN, RF, extreme gradient boost, etc. Leveraging two distinct datasets, one sourced from the renowned UCI ML repository and the additional amalgamated from Kaggle, the research scrutinizes 303 instances with 14 attributes and 1190 patient records with 11 features, respectively. The amalgamated dataset, a fusion of five popular datasets, facilitates a comprehensive evaluation of ML techniques. Results showcase the highest accuracy of 92% using SVM for the UCI dataset, 94.12% with RF for the Kaggle dataset, and an overall peak accuracy of 93.31% employing RF on the combined dataset. These findings underscore the efficacy of ML algorithms in enhancing heart disease prediction accuracy, paving the way for more effective preventative measures in the medical domain.

Ayatollahi et al [47] findings highlight the utility of data mining algorithms, notably the SVM model, in predicting CAD. The SVM model exhibited superior metrics, including lower Mean Absolute Percentage Error, a heightened Hosmer-Lemeshow test result (16.71), and superior sensitivity at 92.23%. Notably, the variables influencing CAD demonstrated a superior gosh of fit in the SVM classifier compared to the ANN model. The SVM algorithm's ROC curve further affirmed its heightened accuracy over ANN. The study advocates for continued research to compare diverse algorithms, aiming to identify the optimal model for disease prediction. The SVM algorithm emerges as a frontrunner, showcasing superior accuracy, performance, and classification capabilities in predicting 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 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 coronary heart disease. This is consistent with other medical research that shows premenopausal women had reduced incidences of subject disease, that is linked to the cardioprotective properties of estrogen. The study also delves into the complex relationship between iron deficiency in younger women due to menstruation, suggesting a potential role in delaying the onset of cardiovascular disease. Rule mining, as a computational intelligence tool, sheds light on essential insights, identifying factors like asymptomatic chest pain and exercise-induced angina as indicative of heart disease for both genders. However, gender-specific distinctions emerge, with resting ECG status identified as a crucial factor for subject disease prediction in females. In contrast, men exhibit a singular rule associating hyper-resting ECG with increased risk. When men and women's healthy indicators are compared, they show similarities, an upward slope, and an oldpeak of less than or equal to 0.56, which are all signs of good health. Beyond the realm of empirical results, this study highlights the value of rule mining and artificial intelligence in elucidating illness variables and navigating the intricacies and contradictions frequently found in medical literature. It is a nuanced exploration that emphasizes the intersection of computational methodologies and gender diversity in understanding the intricate tapestry of heart disease factors.

The study of Leila Baccour [58] begins with the development of the Amended fused TOPSIS-VIKOR methods for classification (ATOVIC), which cleverly integrates the Multiple-criteria Decision Making (MCDM) techniques of VIKOR and TOPSIS into the classification process. Departing from conventional MCDM norms, ATOVIC ventures into uncharted territory, embracing three sets: classes, objects, and attributes (features). Here, criteria metamorphose into features, and alternatives take on the guise of objects tethered to specific classes. A rigorous trial ensues on the CLEVELAND dataset, casting ATOVIC as a luminary in predicting heart disease. Its prowess extends beyond binary and multi-classification scenarios, proving its adaptability in diverse contexts. A symphony of success unfolds as ATOVIC emerges triumphant in forecasting thyroid diseases, eclipsing established classifiers. The algorithm's expedition across datasets (chess, nursery, and titanic) paints a consistent portrait of dominance. ATOVIC's allure lies not only in its accuracy but also in its malleability, allowing seamless parameter adjustments to harmonize with varied datasets. The study takes an intriguing turn toward the future, where the promise of integrating type-1 fuzzy logic and its nuanced iterations (intuitionistic fuzzy logic and type-2 fuzzy logic) beckons. 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.

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 robust features extend to performance evaluation test methods, affording users a panoramic view of system performance. Medical advancement is demonstrated by the development of carotid artery stenting (CAS) as the primary treatment for cerebrovascular stenosis. Yet, this promising avenue is not without its challenges, particularly for older patients who face the specter of Major Adverse Cardiovascular Events (MACE).

In response, Cheng and Chiu [63] embarks on a quest to construct an ANN model, a digital sentinel tasked with evaluating the prognosis of CAS. Armed with data from three hundred and seventeen (317) patients culled from the Taiwan National Health Insurance Research Database (TNHIRD) [64], the ANN model undergoes meticulous training and testing. A constellation of thirteen (13) clinical risk factors forms the input features, while the output corresponds to the incidence of MACE. The ANN model that is produced as a result is an MLP with eighteen neurons in the hidden layer that functions as a digital prognosticator. A performance review demonstrates its efficacy, displaying an astounding 89.4% sensitivity, 57.4% specificity, and 82.5% overall accuracy in the testing group. Widening the lens to encompass the broader patient cohort, the model maintains its efficacy, showcasing sensitivity at 85.8%, specificity at 60.8%, and an accuracy of 80.76%. Beyond its numerical achievements, this ANN model transcends mere prediction. It holds the potential to identify high-risk CAS patients, offering a valuable compass for communication between neurologists and cardiologists in the intricate realm of patient referrals and treatment decisions.

The Back Propagation Neural Network (BPNN) emerges as a particularly effective tool for classifying hypertension gene sequences, demonstrating an impressive 90% accuracy rate for a smaller sample size of 80. After extracting features from specific gene sequences, Zaman and Toufiq [65] present a novel method for classifying hypertension gene sequences using BPNN. They achieve this by using the frequencies of codons, or nucleotide triplets, as a distinguishing metric. Using a range of sample sizes, the study methodically examines how well the BPNN approach performs during the training and testing phases. 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 method demonstrates superior effectiveness in heart disease prediction through the utilization of 14 attributes when compared to alternative approaches. The complexity of heart disease diagnosis necessitates a meticulous examination of patient

clinical test data and health history. Advances in ML, particularly in the development of intelligent automated systems, offer valuable tools for medical practitioners to predict and make informed decisions about diseases. This automated diagnostic system has the potential to enhance timely medical care, leading to life-saving interventions. Meshref [67] presents a comprehensive analysis of the Cleveland heart dataset, employing ML classifiers to optimize diagnostic models. For instance, the MLP model achieved 84.25% accuracy but with an 8-feature set, raising concerns about its appropriateness for subject disease detection. 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.

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

Ref	Method	Dataset	Accuracy %
[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
[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 study, the performance metrics Mean Magnitude of the Relative Error (MMRE) and Logarithmic Standard Deviation (LSD) behave somewhat in opposition to one another, which affects the models selected based on the desired performance 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	Ref	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 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 limitations in dataset scope and size, advocating for self-learning models with minimal data requirements. The research underscores the importance of interpretability and trustworthiness in decision models, advocating for ensemble classification models within the XAI framework. The evaluation of these models using various metrics reveals the robust performance of XAI-driven SVM, LR, and NB, achieving an accuracy of 89%, establishing them as compelling alternatives in comparison to existing models.

In this study [92], the focus lies on the comparative analysis of multiple ML algorithms for predicting heart attack rates based on various features. The investigation delves into determining the importance of these features and assessing the prediction accuracy of the algorithms. Notably, the XGB Classifier emerges as a standout performer, achieving a significant accuracy rate of 86.885%. In the realm of ML, the balance between model complexity, explainability, and prediction performance is crucial. The challenge arises when dealing with intricate algorithms, leading to difficulties in model explainability, a concern addressed by the concept of the black box. To overcome this, XAI methods such as SHAP and LIME prove optimal. SHAP provides a visual representation of model predictions, demonstrated here through its application to the KNN ML algorithm, offering insights into feature importance. Furthermore, LIME is employed to elucidate the workings of the black box model in the KNN algorithm, providing faithful explanations for predictions. This study not only contributes to understanding heart attack prediction using diverse ML algorithms but also showcases the effectiveness of SHAP and LIME in explaining the accuracy of the KNN model. Nascita et al [93] harnessed the power of XAI to clarify, enhance, and implement multimodal DL approaches for addressing various Traffic Classification (TC) tasks, with a specific focus on encrypted traffic. The endeavor involved the design, implementation, and evaluation of an evolved multimodal multitask DL traffic classifier, iteratively refined through distinct stages guided by XAI principles. 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.

Table 2.4: Summary of 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 Cardiac Prediction 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 inclusion of hematological parameters such as “White-Blood-Cells”, “Hemoglobin” and “Platelet-count” adds a layer of granularity, while metabolic variables like “Cholesterol”, “Glucose” and “Triglycerides” contribute to a nuanced understanding of the biochemical milieu. Crucially, the dataset culminates in the pivotal “CoronaryHeartDisease” target class, serving as the focal point for predictive modeling endeavors.

3.1.1 Correlation Matrix and Heat Map

By encapsulating the intricate interplay of diverse variables in the context of cardiovascular dynamics, this dataset stands as a valuable scientific resource poised to decipher the complexities surrounding the manifestation of coronary heart disease. The elucidation of interrelationships among the myriad features within the dataset is facilitated through the construction of a correlation matrix shown in Figure 3.1Error! Reference source not found., a fundamental analytical tool in the realm of data science.

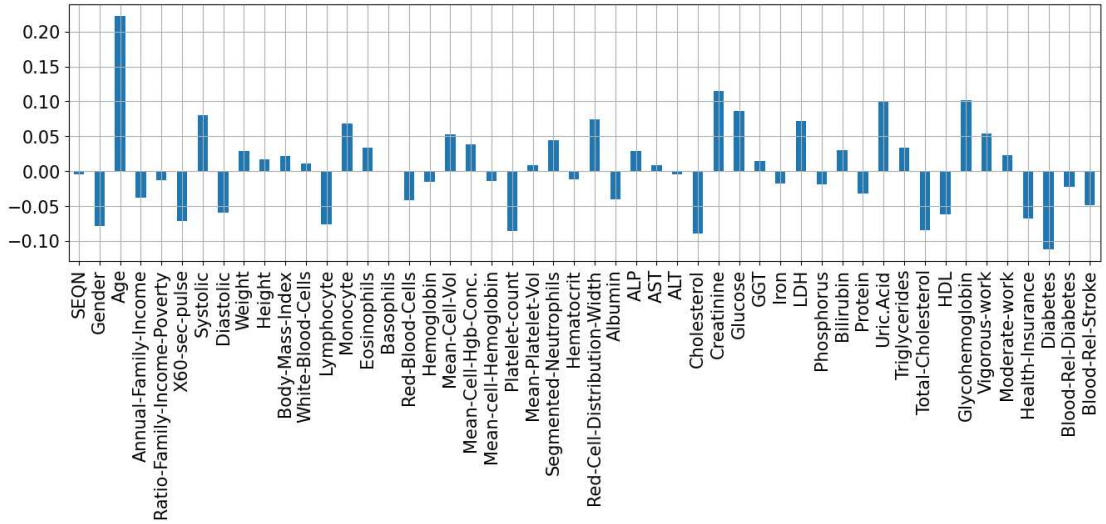

Figure 3.1: Correlation Matrix of Features

This matrix, derived from the XLS file housing 37,079 records and spanning 30 diverse features, encapsulates the pairwise correlations between each variable. A correlation coefficient, ranging from -1 to 1, is assigned to each feature pair, signifying the strength and direction of their linear association. Strong positive correlation is implied by a value near 1, and strong negative correlation is indicated by a coefficient near -1. Such a dimension of correlation matrix acts as a compass for navigating and interpreting complex relationships and patterns between the features of the dataset. Heat map of features is displayed in Figure 3.2. Insights derived from this matrix not only contribute to a nuanced comprehension of the dataset’s internal dynamics but also pave the way for informed feature selection and subsequent model refinement.

3.1.2 Gender Distribution

The dataset under examination offers a noteworthy panorama of gender distribution, a pivotal demographic attribute shaping the landscape of the 37,079 records. A meticulous analysis reveals a nuanced equilibrium, with 51% of the cohort identified as male and 49% as female. This near-equal representation underscores the dataset's commitment to inclusivity, providing a balanced reflection of both genders within its expansive repository. Such gender parity is instrumental in ensuring the robustness and representativeness of subsequent

analyses and predictive modeling endeavors, offering a comprehensive lens through which to explore the intricate relationships between physiological and socio-demographic variables.

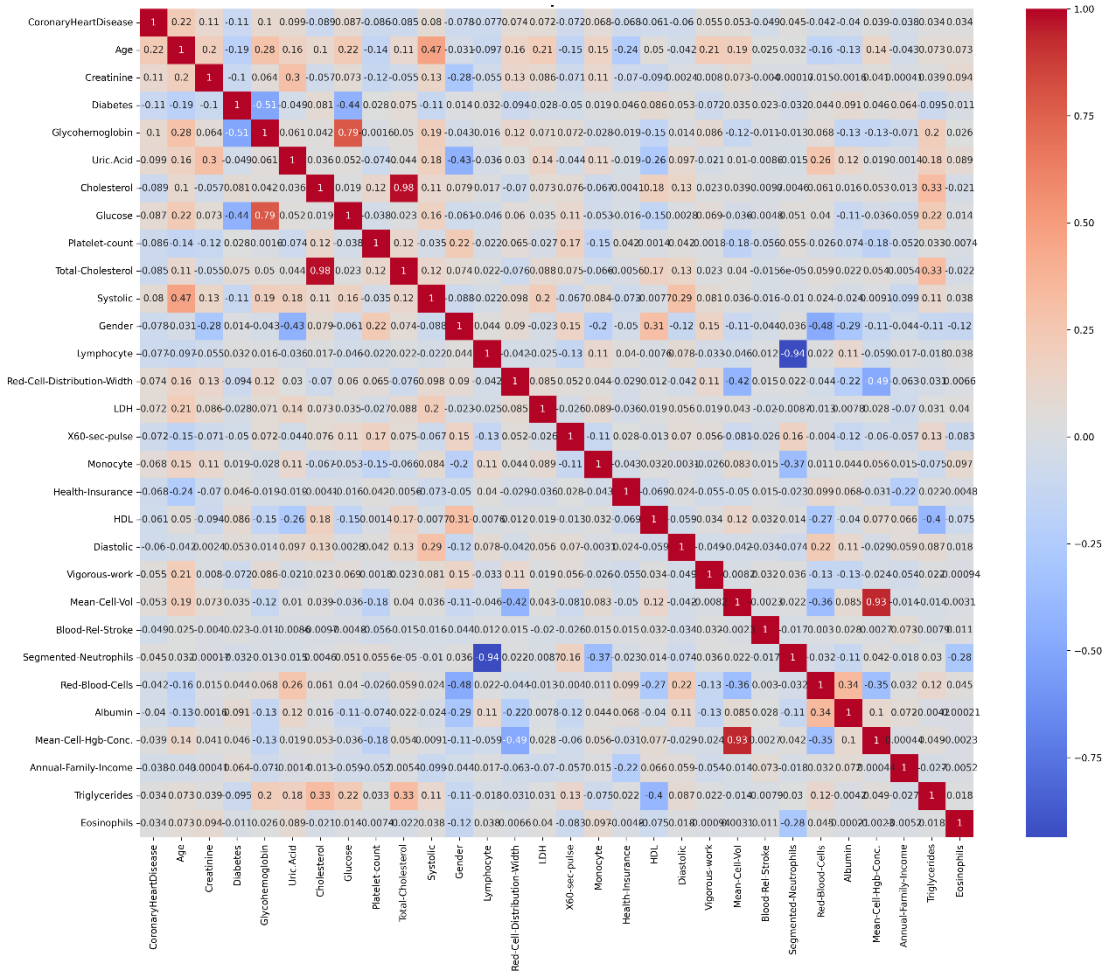

Figure 3.2: Heatmap of Features

3.1.3 Age-Base Distribution

The dataset's exploration into the age-wise distribution of its 37,079 records unveils a nuanced perspective, reflecting a discerning categorization across various age cohorts Figure 3.3. A focal point of interest lies in the age range spanning from 30 to 75 years, where a meticulous examination reveals a density of 0.03%. This particular age group encapsulates a significant portion of the dataset, portraying a diverse representation of individuals within this critical span of adulthood. Furthermore, the dataset extends its scrutiny to those aged between 75 and 88 years, where a heightened density of 0.07% is observed. This demographic granularity underscores the dataset's commitment to encapsulating the intricacies of diverse age profiles, facilitating a comprehensive exploration of health and physiological variables across the lifespan. This stratified approach is paramount in unraveling age-specific patterns and establishing a foundation for robust predictive modeling.

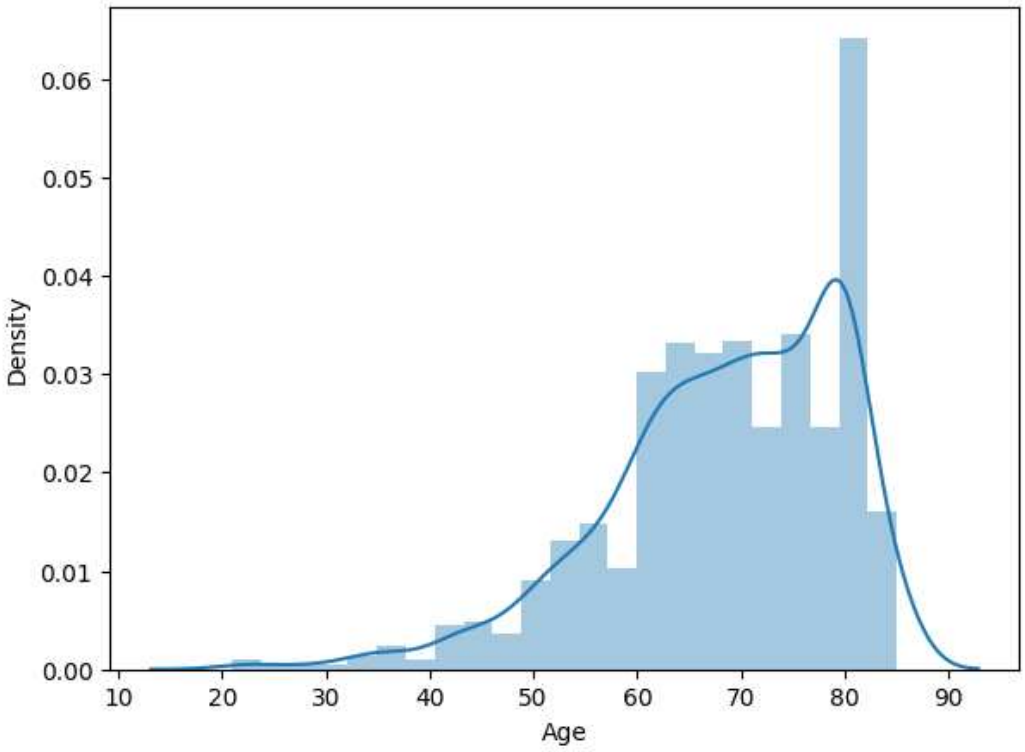

Figure 3.3: Age-based distribution of Dataset

3.1.4 Target Class Distribution and Data Imbalance

The dataset under scrutiny unveils a notable class imbalance, with 35,571 records attributed to individuals categorized as healthy, contrasting with 1,508 records representing individuals diagnosed with the subject disease. This dichotomy translates to a prevalence of 95.93% healthy records and 4.07% records denoting individuals afflicted with cardiac pathology. Such an imbalanced distribution in the dataset can pose challenges during model training, as algorithms may lean towards predicting the majority class, potentially leading to suboptimal performance in capturing minority class patterns. Recognizing the imperative to mitigate this imbalance, study employs the SMOTE from the Python library. This augmentation methodology strategically oversamples the minority class by generating synthetic instances, thereby rectifying the class imbalance and fostering a more equitable representation of both 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 the research dataset, comprised of 51 features for binary classification stored in an XLS file, involves a series of systematic steps executed through Python. The initial phase entails reading the dataset using the Pandas library, followed by a comprehensive examination of its structural attributes, including its shape, column headings, and the count of records. This serves as an essential exploratory step to gain insights into the dataset's dimensions and characteristics. Addressing missing values is a pivotal pre-

processing step to ensure data integrity. The script precisely identifies the existence of missing values, employing a technique that counts the occurrence of null values across the dataset. Subsequently, duplicate records are identified and eliminated, contributing to the refinement of the dataset's quality and preventing redundancy. The subsequent steps focus on the handling of missing values, where the code utilizes a principled approach by dropping records with null values. Furthermore, any residual missing values are replaced with zero to maintain data completeness. The final dataset, after these pre-processing interventions, undergoes shape confirmation to validate the effectiveness of the applied transformations.

Conclusively, the pre-processing regimen encapsulates the resolution of missing values, removal of duplicate records, and imputation strategies, ensuring the dataset's readiness for subsequent modeling endeavors. The utilization of Python, with the Pandas library, underscores a systematic and scientifically rigorous approach to data preparation in the realm of binary classification research.

3.3 Data Balancing and Augmentation

A systematic approach is undertaken to address the inherent class imbalance within the dataset, utilizing the SMOTE or Adaptive Synthetic (ADASYN) techniques. Initially, the script meticulously inspects the dataset's dimensions, emphasizing the pre-augmentation state of the data with 37,079 records and 40 features, with the target variable (CoronaryHeartDisease) comprising binary values denoting the presence or absence of disease. The introduction of SMOTE for the first experiment and for ADASYN technique for the second experiment initiated a transformative process, resulting in a balanced distribution, as evidenced by the equalization of positive and negative class records. This augmentation process, geared towards rectifying the class imbalance, ensures that the predictive model is exposed to a more representative and balanced dataset, fostering enhanced model generalization and robustness in delineating the intricate patterns associated with target class. Subsequently, the dataset undergoes a pivotal partitioning into training and testing sets using the train_test_split method with 80% to 20% ratio respectively, ensuring a stratified split to preserve the original class distribution.

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

The scaling of the features using StandardScaler further standardizes the augmented data, laying a foundation for subsequent model training and evaluation. The meticulous documentation of class distribution metrics throughout the process offers transparency and insight into the effectiveness of the augmentation strategy.

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 see 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.

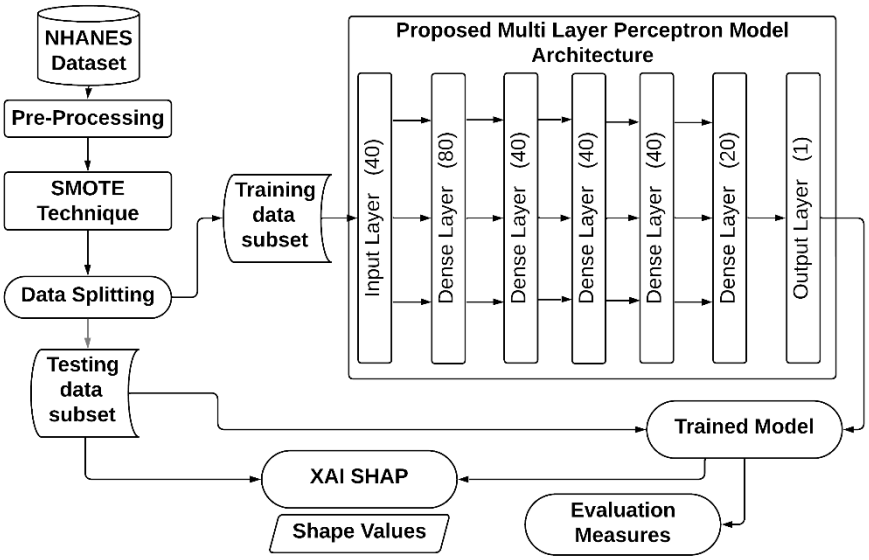

Figure 3.4: Proposed MLP and Attention layer Framework

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 model's interpretability and predictive performance in the context of coronary heart disease prediction.

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 applications like medical diagnosis, where the cost of false positives is significant, precision is especially important. A low rate of false positives is indicated by a high precision score.

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

3.7.3 Sensitivity or Recall

Sensitivity or recall measures how well the model can distinguish true positive cases from false positive instances. It is sometimes referred to as True Positive Rate or Recall. The ratio of True Positives to the total of True Positives and False Negatives is used to compute it. Sensitivity is crucial in scenarios where missing positive instances is highly undesirable, such as in medical screenings.

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

3.7.4 AUC Score and AUROC Curve

As predictive modeling develops, more metrics are used to give a complete picture of a model's effectiveness. AUC Score is a prominent metric used in binary classification problems. Plotting the genuine positive rate versus the false positive rate, it shows the area under the Receiver Operating Characteristic (ROC) curve. Higher discriminating ability is indicated by a higher AUC score. A graphical depiction of a model's performance across various categorization thresholds is called an AUROC (Area Under the Receiver Operating Characteristic) curve. By demonstrating the trade-off between sensitivity and specificity, it provides insightful information about how the model behaves at different decision boundaries.

3.7.5 F1 Score

Recall and precision are balanced by the F1 Score metric. It offers a more comprehensive picture of a model's performance and is the harmonic mean of precision and recall, particularly when working with unbalanced datasets.

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

In summary, the particular goals and features of the issue at hand determine which assessment metrics are most appropriate. Precision, sensitivity, AUC, F1 score, and AUROC curve offer a more detailed knowledge of a predictive model's advantages and disadvantages than accuracy, which just gives a broad picture.

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 software stack comprised essential tools and libraries for ML development. Visual Studio Code (VScode), version 1.85.1, was the integrated programming environment (IDE). Python, the programming language at the core of the ML pipeline, was version 3.10. 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). The use of these tools facilitated efficient data manipulation, analysis, and model development.

4.1.3 CUDA and GPU Configuration

The experimental setup leveraged the capabilities of CUDA with a version of 8700. The system featured a single CUDA device, namely the NVIDIA GeForce GTX 1650, boasting a total memory of 4.29 GB. This configuration was instrumental in harnessing the parallel processing capabilities of the GPU during the model training phase.

4.1.4 Development Libraries and Frameworks

The ML model was developed using TensorFlow and Keras, with versions 2.15.0 for both. These frameworks provided a high-level interface for building and training neural network models. Additionally, the imbalanced-learn library with SMOTE addressed class imbalances in the dataset, enhancing the generalization capacity of the model across various classes.

This carefully configured experimental environment ensured a seamless integration of hardware and software components, creating the ideal environment for the creation and training of a coronary heart disease prediction model. The combination of CUDA-enabled GPU acceleration and a robust software stack laid the foundation for conducting experiments with precision and efficiency.

4.2 Results and Discussion

The proposed MLP architecture, fortified with attention layers, demonstrates exemplary performance across diverse evaluation metrics.

4.2.1 Class Imbalance 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 the training accuracy more than 98% see in Figure 4.1.

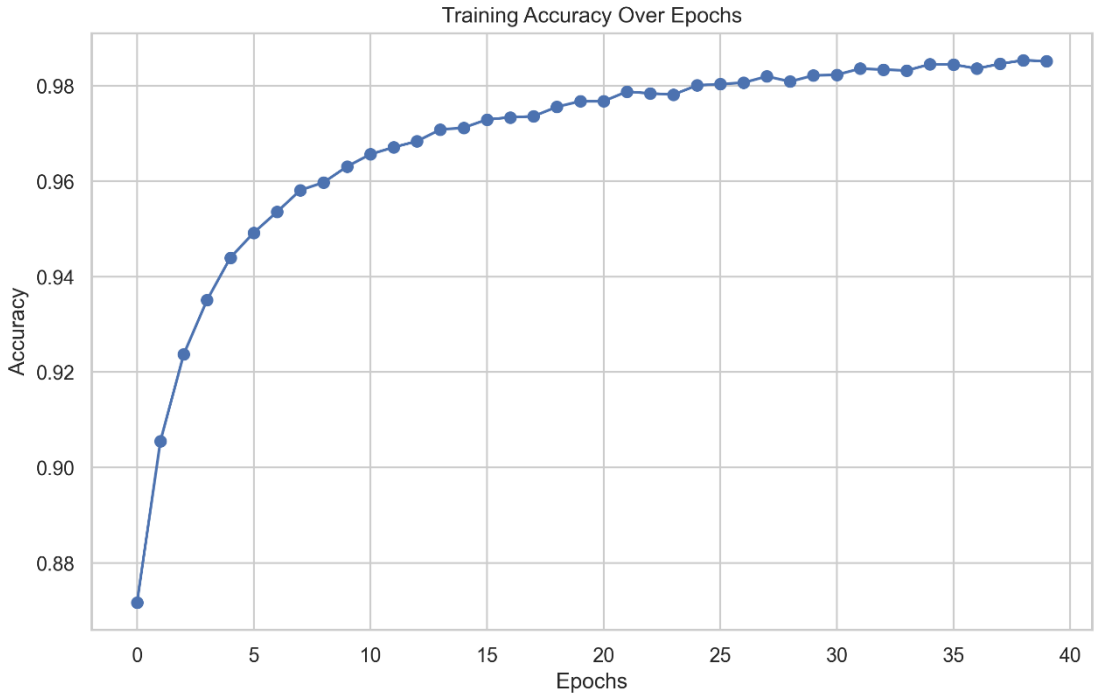

Figure 4.1: Training Accuracy on Class Imbalance Dataset

The model was evaluated on testing dataset and showing a noteworthy accuracy exceeding 95%. There was an enhancement in both recall and precision, with recall reaching 96.71%. The crucial aspect comes to light by plotting the confusion matrix using python lib Science-Kit-Learn (sklearn) and MatPlotLib see in Figure 4.2.

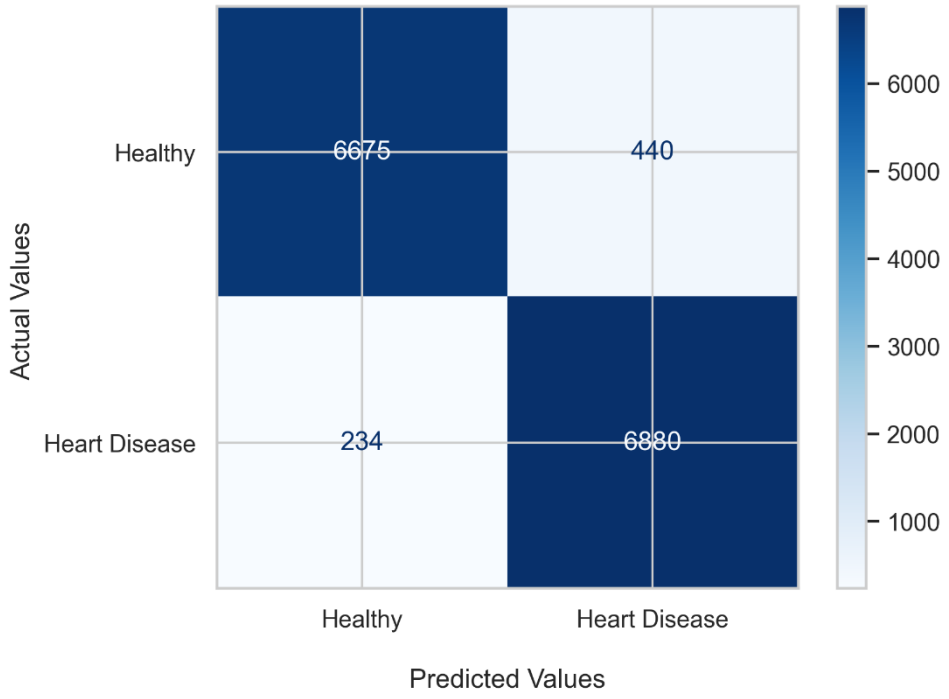

Figure 4.2: Confusion Matrix of Proposed Model on Class Imbalance Dataset

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 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. See in Table 4.1

Table 4.1: Evaluation Measures of Model 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 capacity to distinguish between the two classes is demonstrated by the AUC which is 90.03%. However, this does not adequately account for the difficulties caused by the unequal distribution. In navigating the complexities of imbalanced data, addressing the class imbalance becomes imperative. Additionally, a detailed analysis of specific instances where the model misclassifies positive instances can offer insights into potential areas of refinement. High accuracy is ideal, but evaluating on unbalanced datasets requires a deeper comprehension of metrics than just accuracy. In situations when the cost of missing positive occurrences is large, finding a balance between recall and precision and taking into account the real-world effects of false negatives are critical. Addressing class

imbalance and iteratively refining the model based on its specific application context are key steps toward achieving more robust performance in imbalanced settings.

4.2.2 Balanced Dataset using SMOTE for Data Augmentation

The model's predictive accuracy is encapsulated in the test loss, a metric gauging the dissimilarity between predicted and actual outcomes, registering an impressively low value of 0.0917. This attests to the model's precision in minimizing prediction errors. The test accuracy, a critical indicator of overall classification correctness, attains a remarkable 97.10%, highlighting the model's outstanding level of accuracy in differentiating between target illness cases. The test recall of 97.85% further demonstrates the model's exceptional ability to capture positive case situations. This measure demonstrates how well the model identified a sizable percentage of genuine positive cases in the dataset. Complementing this, the test precision attains an impressive 96.40%, accentuating the model's precision in correctly classifying instances as either positive or negative. See in Figure 4.3 and Table 4.2.

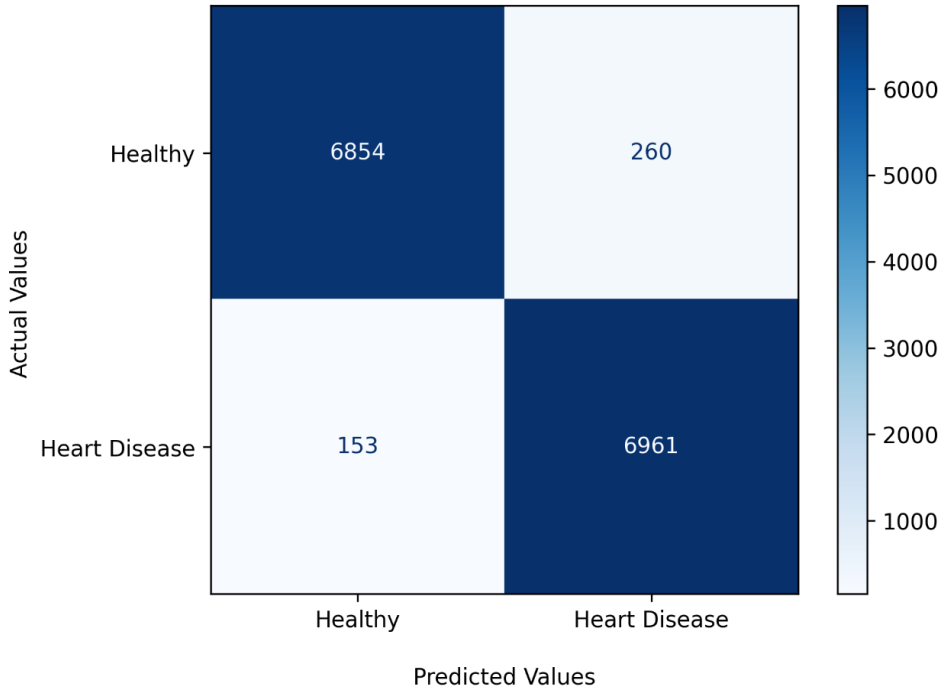

Figure 4.3: Confusion Matrix of Proposed Model with Balanced Dataset

Table 4.2: Evaluation Measures of Model 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

The F1 score of 66.67%, which represents the harmonic mean of precision and recall, shows that the model's overall classification performance is balanced. AUC, which measures the Receiver Operating Characteristic (ROC) curve in Figure 4.4, is a crucial metric for evaluating the discriminatory capacity of the model. With a test AUC of 97%, the model performs exceptionally well in this dimension, demonstrating its ability to discriminate between positive and negative instances with a high degree of accuracy.

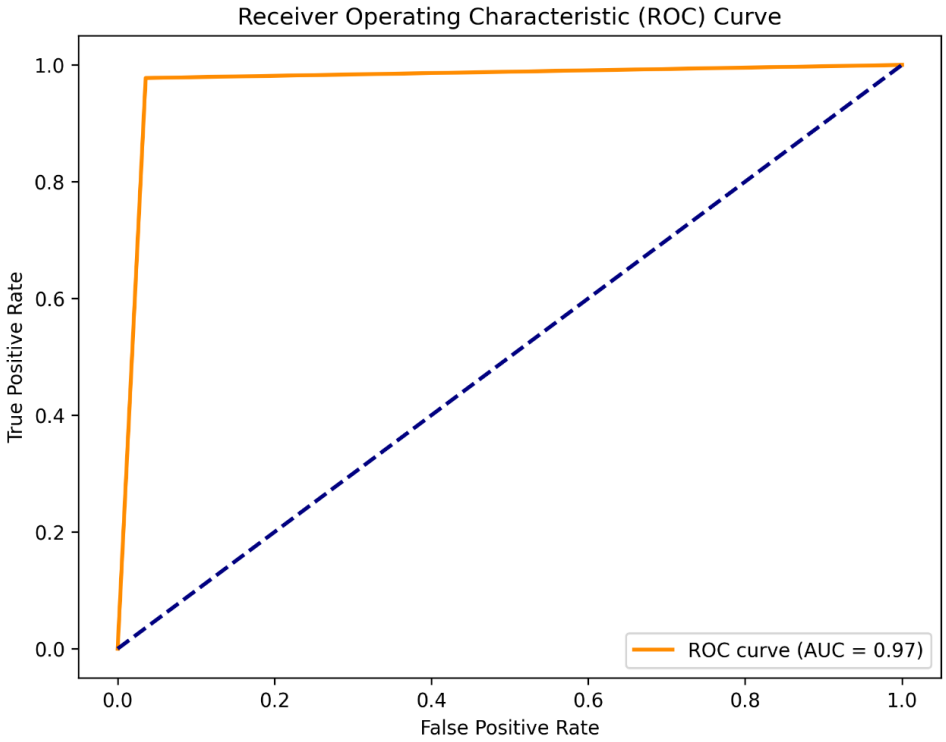

Figure 4.4: ROC Curve of Proposed Model

In summation Table 4.3, the detailed evaluation metrics collectively depict the robust and accurate predictive capacity of the proposed MLP model with attention layers. The results affirm its efficiency in the realm of cardiovascular health prediction, providing a solid foundation for its potential deployment in real-world scenarios with a high degree of reliability and precision.

Table 4.3: Comparison of Both Experiments

S/N	Measure	Imbalanced Data	Balanced Data (SMOTE)
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 an effective method for attributing the model's prediction to certain attributes, which helps to explain the output of ML models. This framework is a straightforward, model-agnostic technique for defining model behavior and improving ML model accuracy. Additionally, it gives a better understanding of how each attribute affects the expected result. A SHAP explanation is used to identify key traits and explain how different elements affect classification results.

4.3.1 SHAP Plot of Model trained on Imbalanced Dataset

Experiment with an imbalanced dataset and the associated evaluation measures, SHAP values can provide important information about how the model makes decisions. The model, trained on the imbalanced dataset, exhibits notable performance characteristics as indicated by a number of assessment measures including the confusion matrix. The confusion matrix reveals that while the overall accuracy is high (95.9951%), the recall for the positive class is considerably low (2.98%). This suggests that the model has difficulty accurately identifying members of the minority class, which could result in a large number of false negatives. Although the precision, at 69.23%, is quite high, it indicates that the model is likely to be correct when it predicts the positive class; nonetheless, the low recall means that a significant number of positive occurrences are being missed. Using SHAP in this case now enables feature-by-feature interpretation of the model's decisions. In order to determine which features contribute most to a positive or negative prediction, SHAP values measure the effect of each feature on the model’s output see in Figure 4.5 and Figure 4.6. Understanding feature importance is crucial in the context of imbalanced datasets, as it helps discern the factors that influence the model’s struggle to identify the minority class.

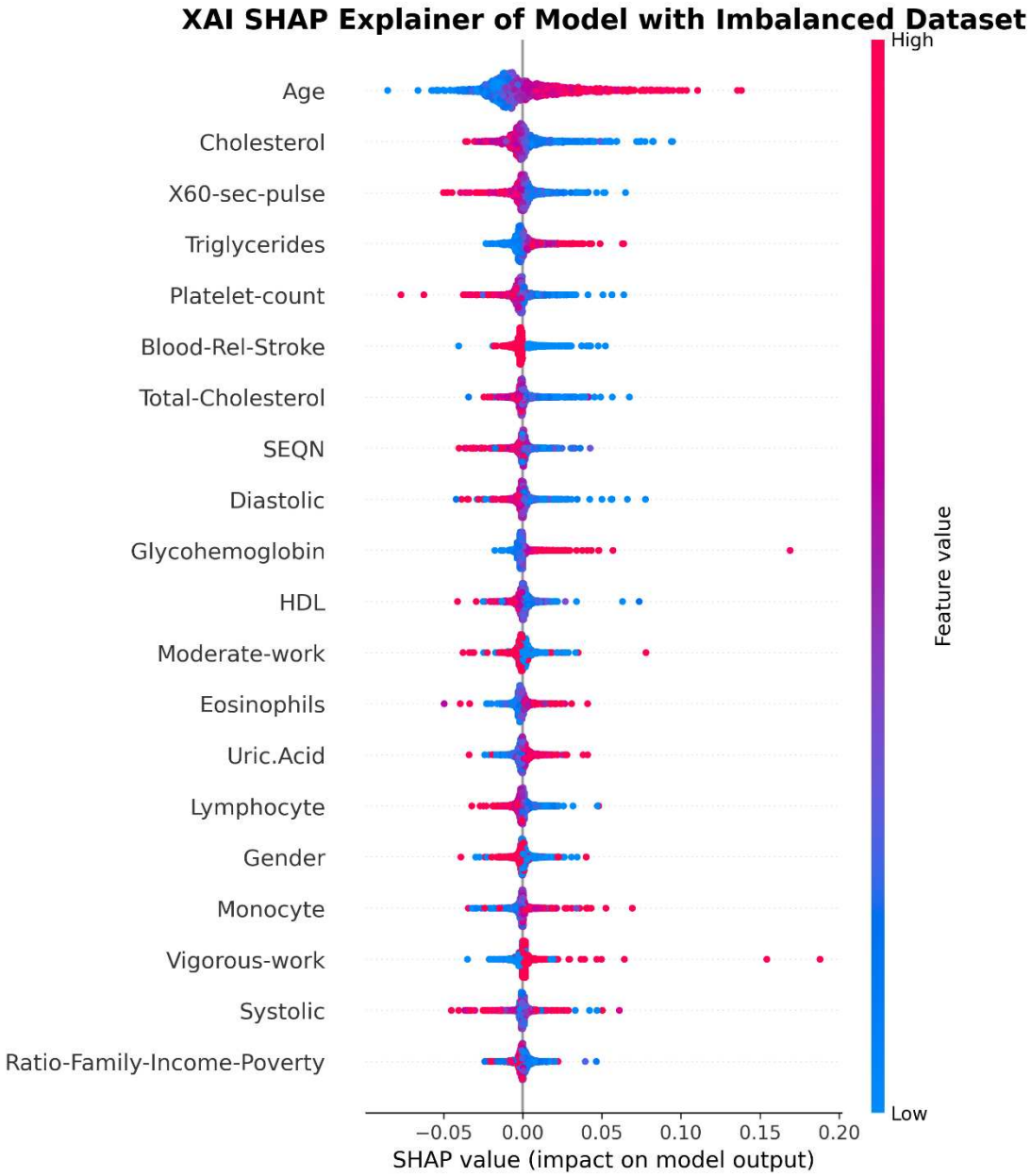

Figure 4.5: SHAP Explainer Plot of Model on Imbalanced Dataset

For instance, SHAP analysis might reveal that certain features play a more significant role in contributing to false negatives. This information can guide further model refinement, such as feature engineering or targeted data augmentation for the minority class. It also allows for a more nuanced understanding of the trade-offs between precision and recall, offering insights into how different features contribute to the model's predictions.

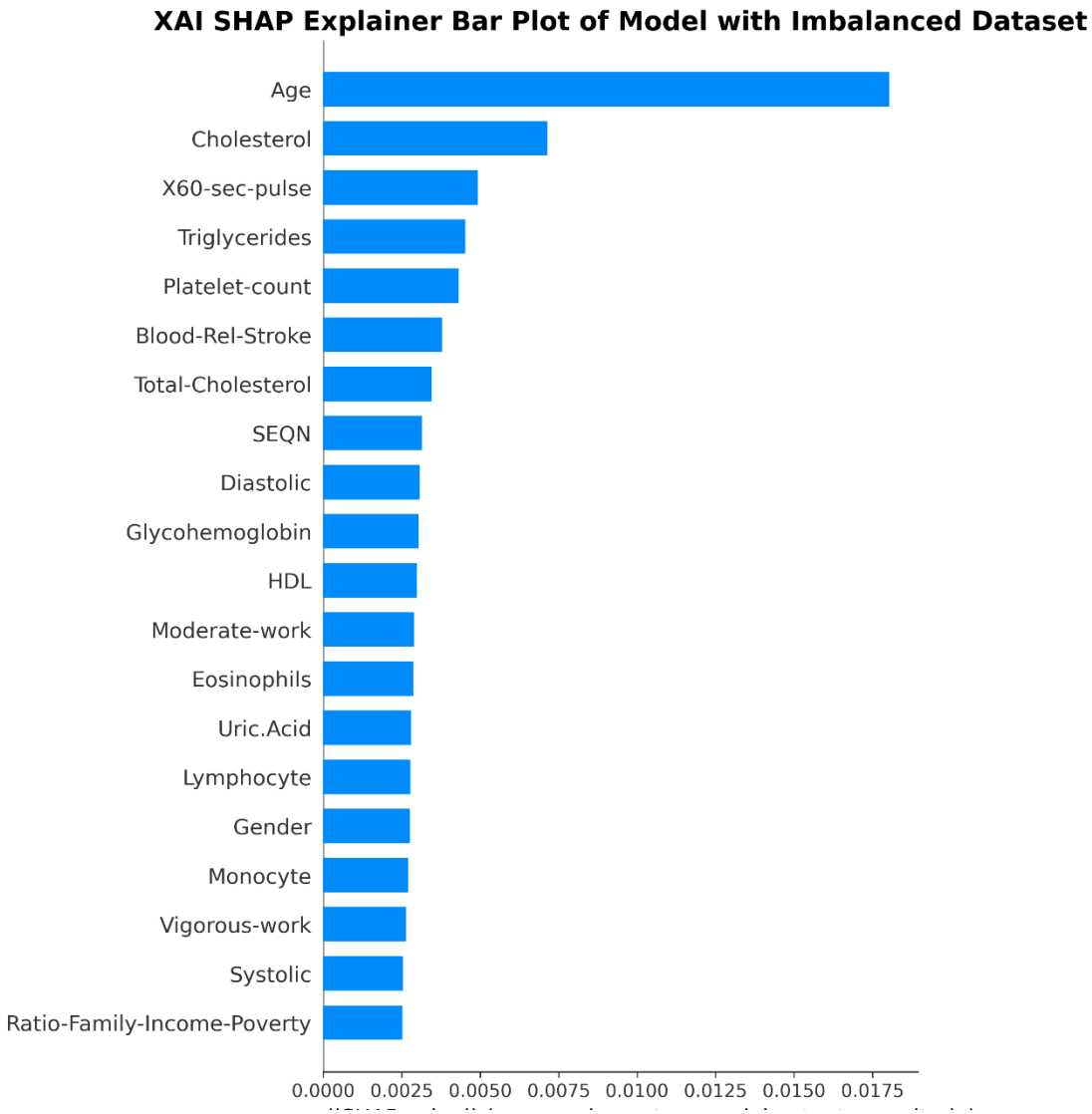

Figure 4.6: SHAP Explainer Bar Plot of Model on Imbalanced Dataset

Applying SHAP algorithms to the model trained on an imbalanced dataset provides a detailed and interpretable view of feature importance. This information goes beyond traditional evaluation metrics, offering insights that can be used to refine the model, address imbalances, and improve overall performance in specific classes. The combination of SHAP analysis and traditional evaluation metrics provides a comprehensive approach to understanding, interpreting, and enhancing the performance of models in complex, imbalanced scenarios. Also see the water-fall plot of individual samples in Figure 4.7

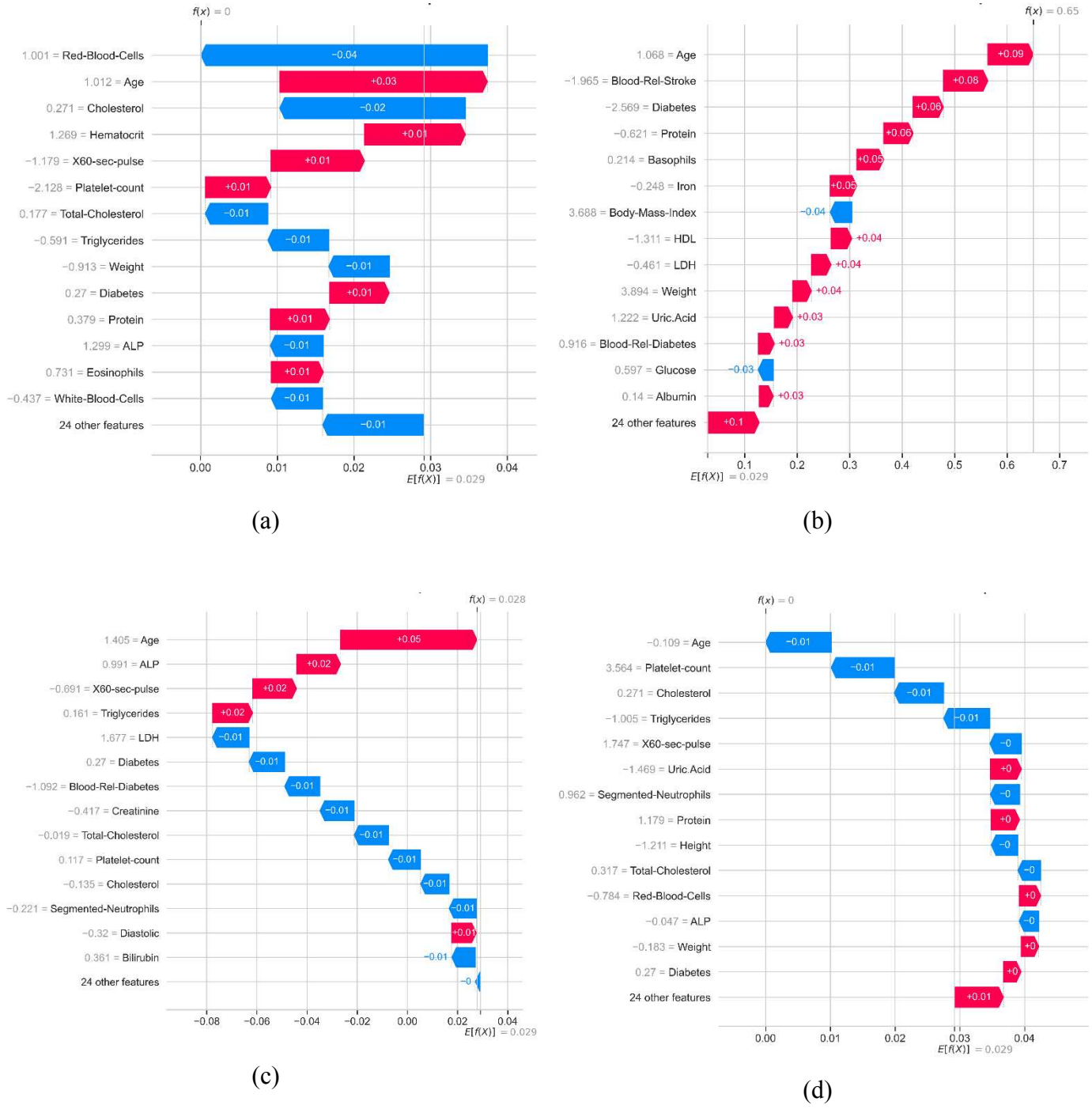

Figure 4.7: SHAP Explainer Water-Fall Plots

4.3.2 SHAP Plot of Model trained on Balanced Dataset

SHAP values were employed to illuminate the predictive dynamics of the proposed DL model concerning the target class. The sequential presentation of feature impact, commencing with “Gender” and progressing through variables such as “Age”, and “Cholesterol” in the SHAP plot, elucidates the discernible influence of each feature on the model’s SHAP Summary plot seen in Figure 4.8 and Figure 4.9.

Table 4.4: Top features from SHAP technique (SMOTE Balanced Dataset)

S/N	Feature Name	SHAP Value (average)
1	Age	0.03381
2	Gender	0.02958
3	Blood-Rel-Stroke	0.02078
4	Triglycerides	0.01794
5	Blood-Rel-Diabetes	0.01712
6	Cholesterol	0.01316
7	Platelet-count	0.01302
8	Diabetes	0.01291
9	Albumin	0.01267
10	Hemoglobin	0.01188
11	Moderate-work	0.01086
12	Diastolic	0.01039
13	Protein	0.00938
14	Height	0.00916
15	X60-sec-pulse	0.00897
16	White-Blood-Cells	0.00879
17	Bilirubin	0.00877
18	Hematocrit	0.00862
19	HDL	0.00856
20	Systolic	0.00831

Upon dissecting the outcomes Table 4.4, discernible patterns emerge, elucidating the pivotal roles of certain features in model predictions. Specifically, “Cholesterol”, “Diabetes”, and “Triglycerides” manifest as substantial determinants, aligning seamlessly with established cardiovascular risk factors within the medical literature. Furthermore, the analysis unveils unexpected contributors for instance, features like “Hemoglobin” and “Red Blood Cells” hint at a potential correlation between hematological parameters and predictions related to coronary heart disease.

Figure 4.8: SHAP Explainer Plot of Model on Balanced Dataset

The implications of these findings extend beyond mere model interpretability, offering practical applications within the medical domain. In addition to adding to the model's credibility, the openness that SHAP values offer enables a clear explanation of the elements influencing predictions. Such insights hold relevance for medical practitioners and researchers, potentially influencing clinical interventions or prompting lifestyle adjustments. Prospective endeavors may involve model refinement based on the recognized feature contributions or delving into supplementary data sources for a more exhaustive analysis.

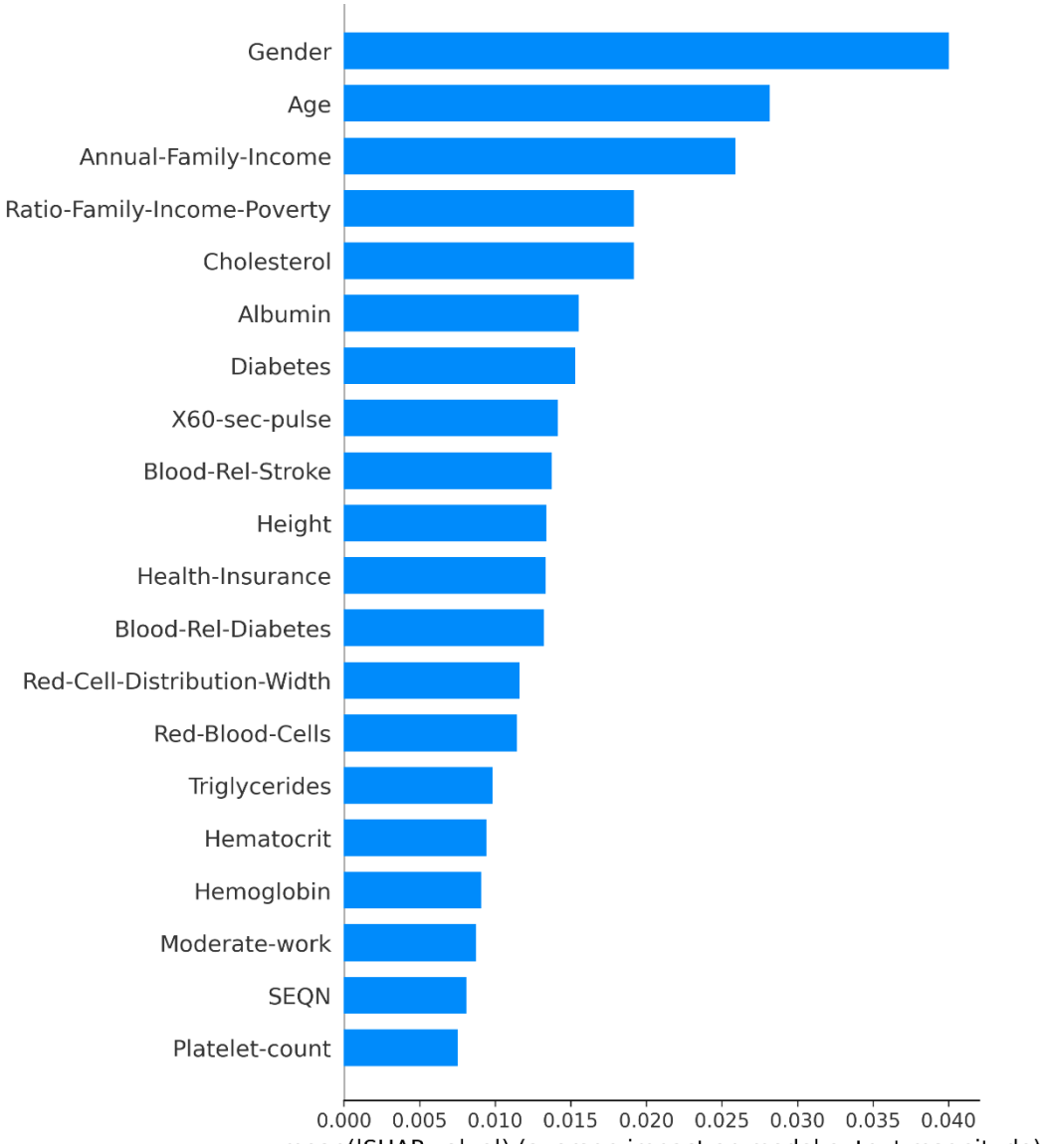

Figure 4.9: SHAP Explainer Bar Plot of Model on Balanced Dataset

The utilization of SHAP values in this context not only amplifies model transparency but also engenders opportunities for ongoing enhancement and a more profound comprehension of intricate feature-prediction relationships within the realm of coronary heart disease. Figure 4.10 shows the SHAP water-fall plot of different individual samples.

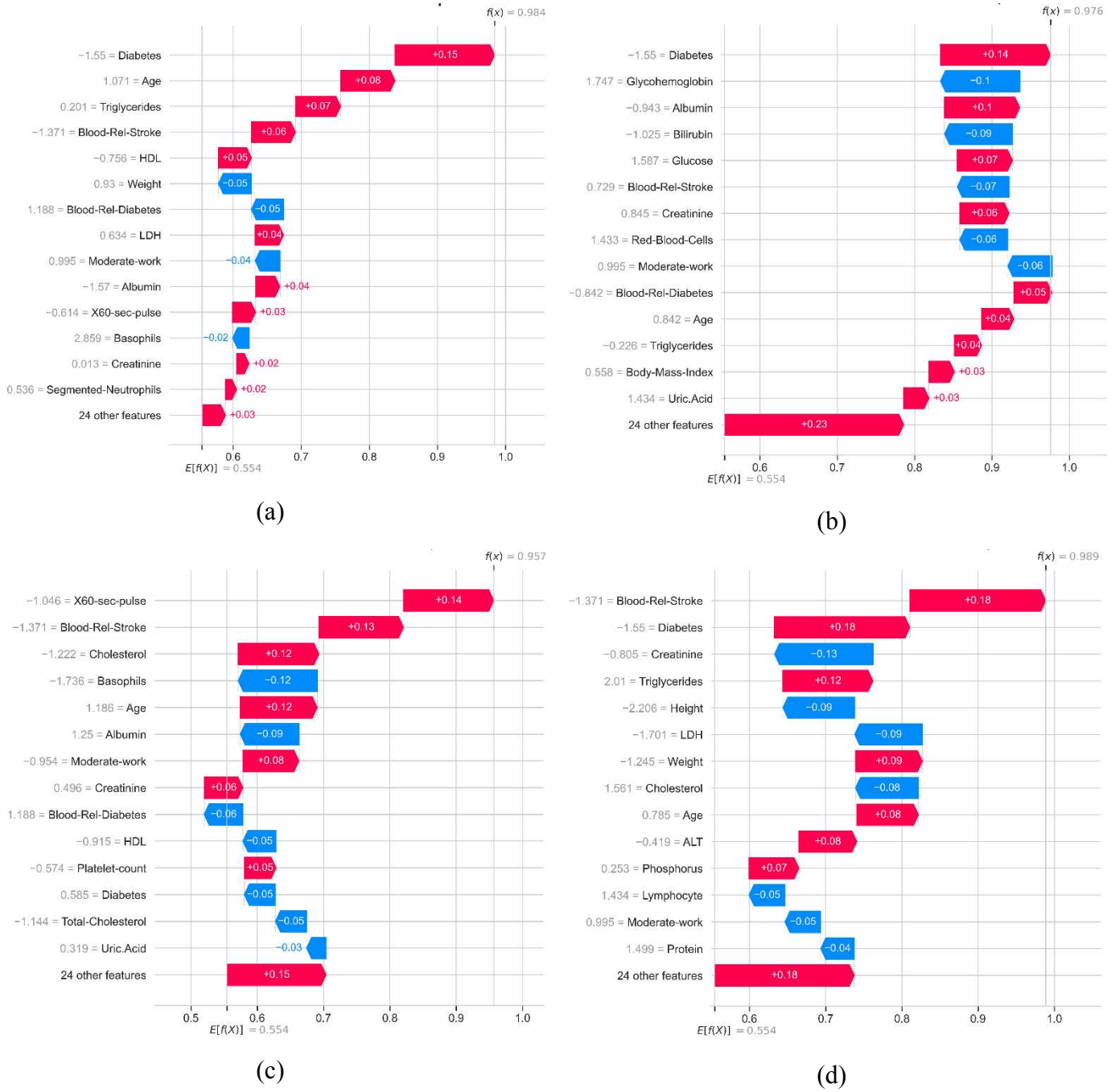

Figure 4.10: SHAP Explainer Water-Fall Plots

Chapter 5: Conclusion and Future Work

In conclusion, this study has endeavored to advance the domain of cardiovascular health prediction by leveraging a sophisticated MLP model enriched with attention layers. The achieved results underscore the model's efficacy, showcasing commendable accuracy, sensitivity, precision, and discriminatory power in identifying instances of coronary heart disease. However, the pursuit of refining predictive models for cardiovascular health necessitates a continued exploration of attention mechanisms in conjunction with various neural network architectures.

The next phase of research should delve into the integration of attention layers with CNNs, RNNs, and LSTMs. CNNs excel in extracting hierarchical features from spatial data, making them particularly suitable for medical datasets. RNNs, with their sequential processing capabilities, are adept at capturing temporal dependencies, crucial for time-series health data. LSTMs, an evolution of RNNs, address the vanishing gradient problem and exhibit superior memory retention, making them valuable for prolonged health monitoring. Exploring the synergies between attention layers and these neural network architectures can potentially improve the model's capability to discern intricate patterns in diverse medical datasets.

The trajectory of future work encompasses a comprehensive exploration of alternative datasets, broadening the model's scope and generalization capabilities. Datasets like Statlog, Cleveland, Hungarian Heart Disease, and Framingham [82] provide unique characteristics and challenges, enabling a more thorough evaluation of the model's robustness across various patient cohorts. These datasets bring forth variations in demographic profiles, risk factors, and healthcare practices, providing a chance to confirm the model's flexibility in practical situations.

In the third phase of research, the focus shifts to enhancing model interpretability through XAI techniques. DeepSHAP [94], DeepLIFT [95], Local Interpretable Model-agnostic Explanations (LIME) [96], FairML, and Causal Explanations (CXplain) [97] stand as pillars in the realm of XAI. A thorough understanding of feature attributions is provided by DeepSHAP and DeepLIFT, which clarify how each feature contributes to model predictions. CXplain offers insights into complex model decision boundaries, enabling a nuanced understanding of intricate patterns. LIME, a model-agnostic technique, facilitates interpretable insights by perturbing instances and analyzing model responses. FairML techniques ensure ethical and unbiased model behavior, imperative in healthcare applications. In practical terms, the integration of these XAI methodologies will involve a meticulous analysis of the proposed model's decision boundaries, feature contributions, and ethical considerations. The transparency offered by these techniques aligns with healthcare standards, ensuring trustworthiness in model predictions. The combination of advanced attention mechanisms, diverse neural network architectures, and comprehensive XAI techniques positions future research endeavors to elevate the predictive capabilities and ethical considerations of cardiovascular health prediction models.

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." arXiv, Feb. 05, 2019. 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/homayoonkhadivi/nhanes-datasets
[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.