CLASSIFICATION OF CARDIOVASCULAR DISEASES (CVDs) USING EXPLAINABLE AI (XAI) BASED ON PHONOCARDIOGRAM (PCG) SIGNALS

By

MUHAMMAD TAHIR JAVAID

NATIONAL UNIVERSITY OF MODERN LANGUAGES,

ISLAMABAD

December, 2024

Classification of Cardiovascular Diseases (CVDs) using Explainable AI (XAI) based on Phonocardiogram (PCG) Signals

By

Muhammad Tahir Javaid

BEEE, HIET, Islamabad 2013

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In **Electrical Engineering**

TO

FACULTY OF ENGINEERING AND COMPUTING

NATIONAL UNIVERSITY OF MODERN LANGUAGES ISLAMABAD

THESIS AND DEFENCE APPROVAL FORM

The undersigned certify that they have read the following thesis, examined the defense, are satisfied with overall exam performance, and recommend the thesis to the Faculty of Engineering and Computing for acceptance.

Thesis Title: Classification of Cardiovascular Diseases (CVDs) using Explainable AI (XAI) based on Phonocardiogram (PCG) Signals

Submitted by: Muhammad Tahir Javaid **Registration #:** 13-MS/EE-F21

Master of Science in Electrical Engineering

Electrical Engineering

Dr. Sheraz Alam _____

Research Supervisor — Signature of Research Supervisor

Mr. Muhammad Waqar _____

Research Co-Supervisor — Signature of Co-Supervisor

Dr. Farhan Sohail _____

HOD (EE) — Signature of HOD (EE)

Dr. Muhammad Noman Malik _____

Dean (FEC) — Signature of Dean (FEC)

26th December, 2024

Date

AUTHOR'S DECLARATION

I Muhammad Tahir Javaid

Son of Muhammad Javid.

Registration # 13-MS/EE-F21

Discipline Electrical Engineering

Candidate of **Master of Science in Electrical Engineering (MSEE)** at the National University of Modern Languages does hereby declare that the thesis **Classification of Cardiovascular Diseases (CVDs) using Explainable AI (XAI) based on Phonocardiogram (PCG) Signals** submitted by me in partial fulfillment of MSEE degree, is my original work, and has not been submitted or published earlier. I also solemnly declare that it shall not, in the future, be submitted by me for obtaining any other degree from this or any other university or institution. I also understand that if evidence of plagiarism is found in my thesis/dissertation at any stage, even after the award of a degree, the work may be canceled, and the degree revoked.

Signature of Candidate

Muhammad Tahir Javaid

Name of Candidate

26th December,2024

Date

ABSTRACT

Classification of Cardiovascular Diseases (CVDs) using Explainable AI (XAI) based on Phonocardiogram (PCG) Signals

Cardiovascular diseases (CVDs) are among the leading causes of death worldwide, making early heart examination crucial. Analyzing heart sounds is one of the many key methods for diagnosing cardiac disorders. However, automated classification of heart sounds remains challenging. Phonocardiograms (PCGs) offer a non-invasive method for identifying CVDs by capturing continuous heart sounds, including murmurs. Recent advancements in artificial intelligence (AI) and machine learning (ML) have made it feasible to analyze large volumes of PCG data from cardiac cycles within a reasonable time frame. Researchers have leveraged these technologies in numerous case studies over the past few years to improve detection accuracy and reduce detection time. A comparatively recent shift in this regard is the focus on improving the interpretability and trustworthiness of these AI-driven diagnostic models, a field known as Explainable AI (XAI). XAI is crucial because it not only provides insights into how models make predictions but also fosters trust among clinicians and patients, ensuring that decisions are based on understandable and justifiable reasoning. This transparency is particularly important in healthcare, where the consequences of misinterpretations can be significant. This study focuses on feature extraction, classifier selection, and model interpretability for efficient XAI implementation. Three ML classifiers [Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM)] are used to predict CVD risk. While all ML models demonstrated good prediction capability RF achieved the best performance with an accuracy of 93.82%, precision of 92.01%, recall of 95.33%, specificity of 92.44%, and an F1 score of 93.64%. Besides critical predictors of long-term CVD risk and its impact on risk prediction are obtained using an explainable techniques for interpreting ML predictions.

Keywords: Cardiovascular diseases; Phonocardiograms; Artificial Intelligence; Machine Learning; Explainable AI; Trustworthiness

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CVDs	Cardiovascular Diseases
PCG	Phonocardiography
ECG	Electrocardiography
ML	Machine Learning
DL	Deep Learning
AI	Artificial Intelligence
XAI	Explainable AI
CAD	Computer-aided Diagnosis
WHO	World Health Organization
LMICs	Low- and Middle-Income Countries
CAHSA	Computer-Aided Heart Sound Analysis
SDGs	Sustainable Development Goals
UN	United Nations
NCDs	Noncommunicable Diseases
WHF	World Heart Federation
NGOs	Non-Government organizations
SVMs	Support Vector Machines
HMMs	Hidden Markov Models
HSMMs	Hidden Semi-Markov Models
PSD	Power Spectral Density
MFCC	Mel-Frequency Cepstral Coefficients
SVM	Support Vector Machines
NB	Naive Bayes
DT	Decision Trees
CinC	Computers in Cardiology
TWSVM	Twin Support Vector Machine
MDS	Multidimensional Scaling

KNN	K-nearest neighbors
ANN	Artificial Neural Network
F-NN	Feed-Forward Neural Network
CNN	Convolutional Neural Network
DCA	Deep Convolutional Autoencoder
DLUTHSDB	Dalian University of Technology Heart Sounds Database
STFT	Short-Time Fourier Transform
TFD	Time-Frequency Domain
DNN	Deep Neural Network
MGWST	Modified Gaussian window-based Stockwell Transform
STKE	Shannon-Teager-Kaiser Energy
SE	Shannon Entropy
SAEs	Stacked Autoencoders
RF	Random Forest
CWT	Continuous Wavelet Transform
MLP	Multi-Layer Perceptron
SHAP	Shapley Additive explanations
LIME	Local Interpretable Model-Agnostic Explanations
SMOTE	Synthetic Minority Over-sampling Technique
Std	Standard Deviation
Min	Minimum
Max	Maximum
RMS	Root Mean Square
CM	Confusion Matrix
TP	True positive
TN	True Negative
FP	False Positive
FN	False Negative

ACKNOWLEDGEMENT

Primarily, I would like to give all the standard to Allah Almighty for His blessings on me. Secondly, I am so grateful to my family for their support to me during my research period. They have always supported me morally, encouraged me, and always believed in me and my capabilities which have led me to success.

I also express my gratitude to my research supervisor Dr. Sheraz Alam and co-supervisor Mr. Muhammad Waqar for their continuous hard work and great support in the completion of this research. The constant passion that they have shown and their knowledge of the field of research have not ceased to amaze me I am honored to have obtained their advice, help, and constant participation in my work. Their unique help has been central to the formation of this thesis in question.

DEDICATION

To my wife, the source of strength, for your patience, and support during the preparation of this thesis. Her encouragement in me made every task easier regardless of how hard it may seem to be, it is done step by step.

CHAPTER 1

INTRODUCTION

1.1 Cardiovascular Diseases (CVDs)

Heart and blood vessel disorders are individually known as coronary heart disease, congenital heart disease, minor arterial disease, and as a collective called cardiovascular disease or CVD. Cardiac auscultation, commonly known as listening to the heart acoustics, is one of the processes used in diagnosing some of the cardiac ailments [1]. Most of these diseases are an interaction of lifestyle, the physical environment, and heredity factors including diet, lack of exercise, smoking, and excessive drinking.

Figure 1.1 shows the subdivisions of CVD which include Atrial Fibrillation, Valvular Heart Disorder, Heart Failure, Congenital Heart Disorder, Cardiomyopathy, and CAD. Such representation vividly highlights the versatility of reconciling AI techniques in promoting the detection and handling of various forms of CVDs with enhanced diagnosis and favorable patient experiences.

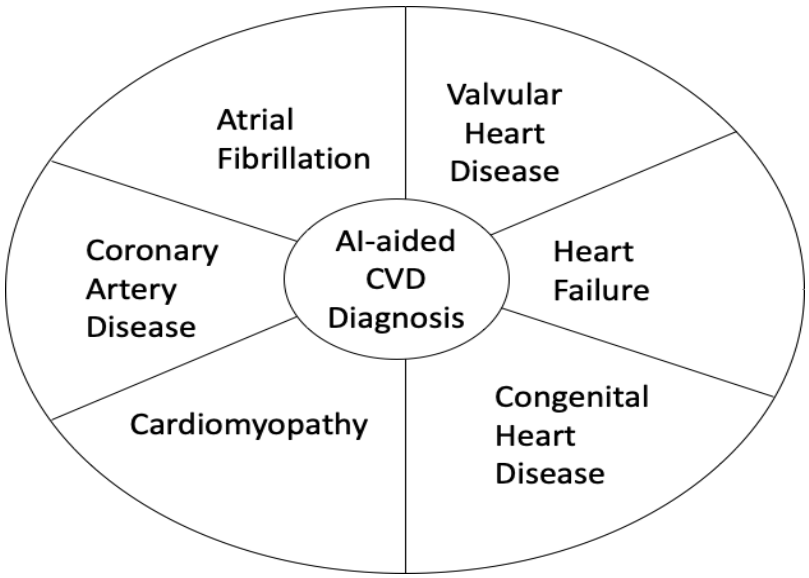

Figure 1.1: Types of CVDs [2]

1.1.1 Global Impact of CVDs

CVDs account for the highest global mortality rates, but they remain significant in terms of diagnostic challenges, especially for resource-constrained facilities. CVDs are known to claim thirty-two percent of all global deaths, estimated to be around 17.9 million deaths taking place annually, to be precise [3]. Relatively in low and middle-income countries (LMICs) medical facilities and professional health care are limited this is why the toll of such diseases is super excessive [4]. Nevertheless, reaching a professional diagnosis is still impossible due to the patient-to-doctor ratios that may reach 50000: 1 in some regions, make a diagnosis through a mobile application or cloud. [5], [6]

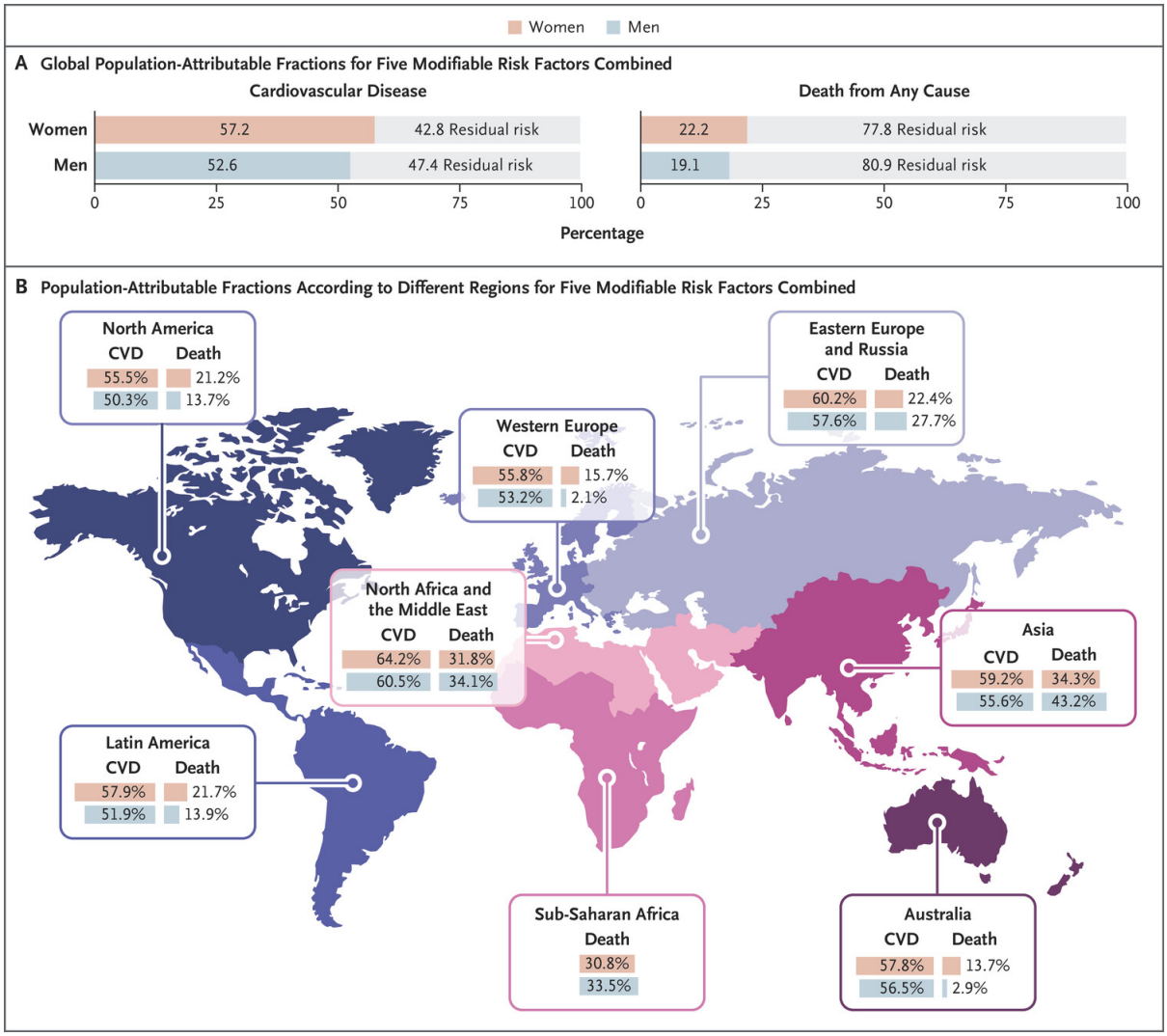

Figure 1.2: Mortality rate of CVD Globally [7]

Figure 1.2 depicts how those risk factors that are modifiable and those that cannot be altered influence the ten-year prevalence of cardiovascular disease (CVD) and all-cause death based on geographical location as well as gender differences. Risk factors that can be changed or controlled are the type of foods eaten, exercise, smoking, and alcohol habits that can be altered in some way or another. In contrast, fixed risk factors which are also known as unchangeable or behavioral risk factors cannot be altered, they include age, gender, and genetic susceptibility. It also represents those well-established modifiable risk factors, as well as independent global and regional effects concerning North African and Middle Eastern nations, North American and Asian countries, Australians, and West Europeans.

1.1.2 Essential Need for Early and Accurate CVD Diagnosis

The heart sounds and murmurs are small in amplitude and frequency signals making them nearly clinically inaudible. In modern clinical practice, doctors apply conventional systems like mechanical stethoscopes to auscultate heart sounds and murmurs. It results in rather low accuracy and, consequently, incorrect diagnoses are sometimes made. Additionally, conventional methods cannot capture the sounds as measured and are, therefore, highly dependent on the doctor's abilities, which inevitably degrades over time. Solving this problem is urgent for the early diagnosis of pathologic changes in heart sounds. [8]

LMIC healthcare centers are generally understaffed and underequipped to offer adequate evaluations for cardiovascular abnormalities, instead, the facilities may even be without a stethoscope. This dependence on auscultation, compounded by the qualitative difference in clinician's experience, frequently results in either a failure to diagnose or a delay in the course of a correct diagnosis, as well as in an enhanced likelihood of adverse outcomes due to the necessary delay in the administration of accurate treatment. Hence, there is a dire need for new approaches that can improve diagnosis in LMICs, possibly by implementing the use of technological diagnostic tools and artificial intelligence to assist clinicians with their diagnostics and overall patient care. CVD Diagnostic methods are depicted in figure 1.3.

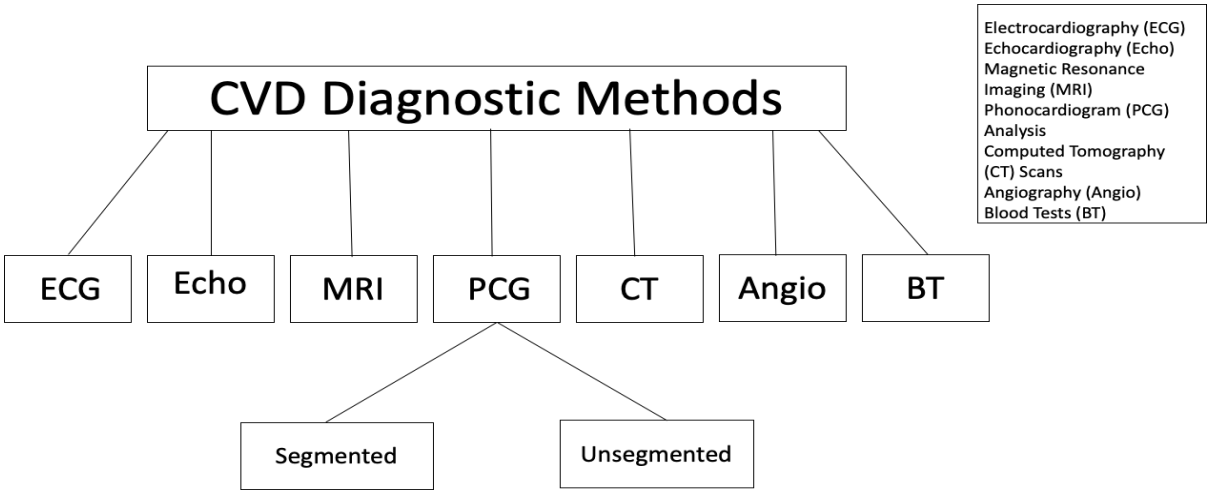

Figure 1.3: CVD diagnostic methods

1.1.3 Traditional and AI-Based CVD Diagnostic Solutions

The solutions for diagnosing CVDs have evolved significantly, moving from traditional clinical practices to advanced AI-based approaches. Traditional solutions like ECGs rely on clinician expertise and can be time-consuming and resource-intensive. In contrast, AI-based solutions use advanced algorithms and large datasets to improve accuracy, efficiency, and personalization. Integrating AI can enhance diagnostic precision, ease clinician workloads, and improve patient outcomes. [9]

Table 1.1 describes the contrast between the conventional and the intelligent techniques for detecting CVDs. Conventional techniques are known methods used often in the clinic but issues like expense, and the requirement of special tools may constrain their usage.

Table 1.1: Comparison of Traditional and AI-Based CVD Diagnostic Solutions

Aspect	Traditional Solutions	AI-Based Solutions
Method Names	ECG, PCG, Clinical Assessments, Echocardiogram, Stress Tests, Cardiac catheterization	ML & DL, AI for Image Analysis, Predictive Analytics
Accuracy	Dependent on clinician expertise and diagnostic tools	High accuracy with large datasets
Speed	Time-consuming	Faster data processing and real-time analysis capabilities
Cost	Often high due to the need for specialized tests and equipment	Initial setup can be costly, but operational costs may be lower
Invasiveness	Some methods are invasive (e.g., cardiac catheterization)	Mostly non-invasive, relying on data analysis
Data Requirements	Relies on individual patient data and clinician judgment	Requires large, high-quality datasets for training and validation
Interpretability	Generally well-understood and accepted in clinical practice	Challenges in interpretability and transparency of AI models

On the other hand, AI-based methods offer several advantages, including higher accuracy and speed, reduced invasiveness, and the potential for more personalized care. ML algorithms and DL models can analyze large datasets to identify patterns and make predictions with high precision, often surpassing human capabilities. However, the implementation of AI in healthcare also presents challenges, such as the need for extensive high-quality data, issues with model interpretability, and the addition of AI tools into clinical workflows.

1.1.4 Advantages of Unsegmented PCG in Clinical Diagnostics

Depending on the type of interface PCG signals can be segmented or unsegmented. Segmented PCG is not favored despite its seemingly systematic representation of the heart sounds because of its impracticality in clinical situations as compared to unsegmented PCG. The segmentation methods commonly use synchronized ECG recordings that can be challenging to get particularly in dealing with newborns or in noisy environments. Sometimes, the signal might be quite noisy and fluctuate from one patient to another, and as a result, techniques such as envelope detection might fail to detect all the essential peaks of the heart sound or identify some of them as false ones. On the other hand, unsegmented PCG does not present with these problems and can thus sustain a steady monitoring of the heart conditions, hence an improved classification of these conditions using ML. Research [10] has revealed that the unsegmented PCG methods are more effective and accurate, which makes the method very useful in the diagnosis of cardiovascular diseases irrespective of groups of patients. The presented research employs unsegmented PCGs to eliminate the mentioned shortcomings.

1.1.5 Explainable AI (XAI)

According to [11], it is the capacity to communicate the decision-making process of artificial intelligence (AI) to a wider range of end users in a way that is both clear and concise. XAI is defined as a collection of methods and approaches. The explainability of the model is of more relevance to data technologists or specialists. Physicians and medical professionals, however, are more focused on clinical inference and prediction. Interpretability is the other concept associated with explainability. The ability to explain an abstract idea is known as interpretability [12]. While interpretability refers to the model representation derived from the training data, explainability deals with the interpretation of predictions produced in the presence of renewed cases. [13]

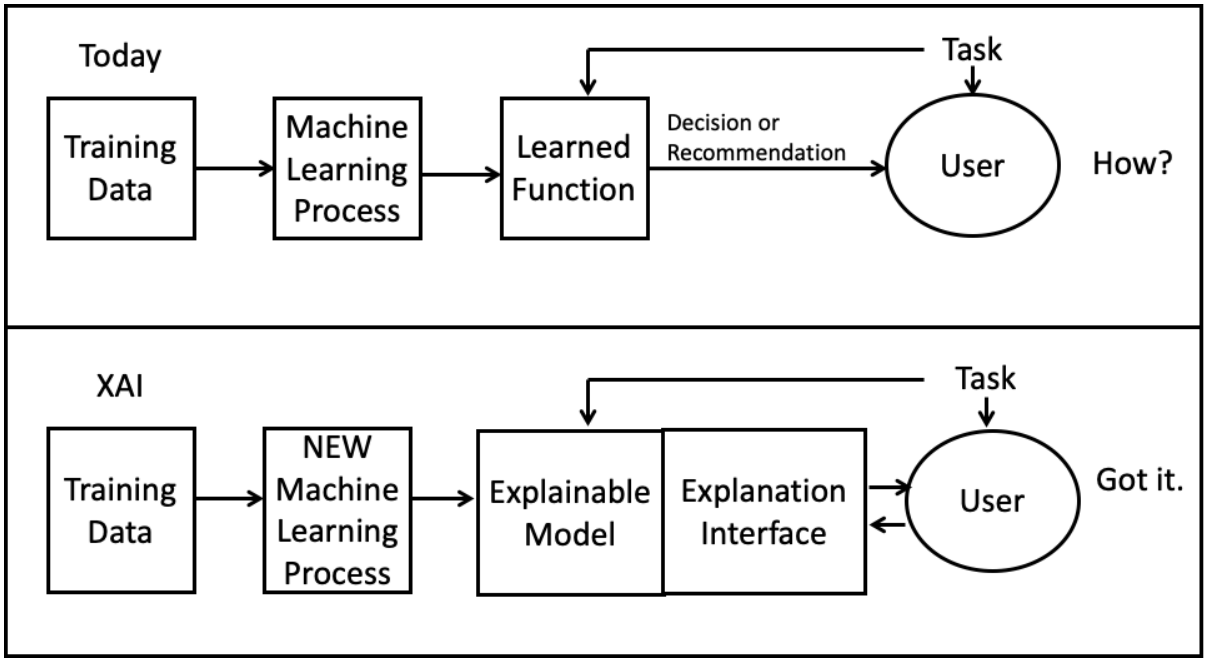

Figure 1.4: Explainable Artificial Intelligence Concept [14]

Figure 1.4 illustrates the distinction between traditional ML processes and XAI processes, emphasizing the need for transparency in AI systems. In today's ML framework, training data is processed through an ML algorithm to produce a learned function, which then provides decisions or recommendations for users. However, this approach often leaves users questioning the rationale behind the outcomes, as there is no clear explanation of how the model arrived at its conclusions.[14]

Whereas, the XAI framework is a new approach in ML that results in an explainable model. This model generates solutions or recommendations but has an explanation layer that explains why it concluded. Thus, the topic of how the conclusions were made is reflected by this extra degree of openness, which is helpful to the users to know the particulars of the process through which AI decisions are made so they can trust the result. The results of XAI are benefits like increased user confidence and its applicability to health care and other sectors due to the provision of brief and less complex explanations of the model's results.

1.1.6 The Importance of XAI in Medical Applications

The value of XAI has gained widespread recognition in business and academics in recent years. The ML and DL model's decision-making processes are difficult to understand due to their high level of complexity. These models are opaque black boxes that generate predictions based on incoming data but do not explain their logic [15]. Traditional ML models may have significant limitations due to their lack of interpretability and transparency, which can result in several issues and difficulties.

Indeed, the difficulty of validating and establishing traditional ML models can be considered one of the major issues. This characteristic provides the models with a certain degree of opacity, which complicates the understanding of how they work and deliver results. Due to such issues, people may not be able to trust and understand how such models work

hence limiting their usage and dependency of such models. The inability of conventional ML models, along with their limitations, and the demand for accurate, explainable, and trustworthy models that are fairly developed and able to address the faults of unfair models are the causes of the XAI need.

In the last few years, XAI solutions have gained prominence in understanding the foundation behind the decision-making procedure of the ML models mainly in the healthcare industry. Well-known techniques of SHAP (Shapley Additive explanations), LIME (Local Interpretable Model-agnostic Explanations), besides other comparable methodologies will be described in the background section of the literature review.

1.2 Motivation

It is observed in the above studies that the increasing prevalence and impact of CVD necessitate improved diagnostic methods. Traditional ECG-based techniques often miss mechanical abnormalities, while PCG offers additional insights into heart health. However, AI-based PCG classification, similar to several other classification techniques, raised the "black box" problem, meaning that little to no one can understand how the model made the decision, which is a severe issue with doctors and patients. This is important for XAI since it develops models that are easily explained and which help in building trust. However, classical stethoscopes, the quality of the used instrument, and human hearing perception are drawbacks of heart auscultation. This research will endeavor to enhance the optimality of up-to-date solutions by making them stronger, more accessible, and easier to employ to improve diagnostic acuity and clinician acceptability.

1.3 Problem Statement

Cardiovascular disease (CVD) is the leading cause of death worldwide, accounting for an estimated 17.9 million deaths in 2019. Early detection and treatment of CVD are essential for improving patient outcomes, but traditional screening methods are less efficient, subjective, and error-prone. A potential solution to this is to provide automated diagnosis [4]. Existing cardiovascular disease (CVD) models are less trustworthy because they are not explainable. This means that we cannot understand how the models make their predictions, which makes it difficult to trust their results.

Moreover, the rapid emergence of new cardiovascular examination and treatment technologies is generating increasing amounts of data and information. This rapid technological advancement has made the work of cardiologists more demanding and highlighted the need for automated screening techniques that can provide cost-effective healthcare solutions without compromising patient well-being.

1.4 Aim and Objectives

Aim: To develop an XAI-based model for the classification of CVDs that can provide precise predictions while also being interpretable.

Objectives: The primary objectives of this research are as follows:

- To implement existing machine learning models for the automated classification of cardiovascular diseases from PCG data.
- To integrate state-of-the-art explainability techniques into the AI models to generate interpretable results.
- To evaluate the model's performance in terms of accuracy, interpretability, and clinical relevance.

1.5 Scope of Research

CVDs are truly a massive family of disorders that embrace mechanical and vascular diseases. This proposed research emphases on the use of XAI for the classification of these diseases using unsegmented PCG signals. It also includes building AI models that are explainable and capable of estimating the level of disease severity and distinguishing between various Cardiovascular diseases using the PCG data to improve diagnostic performance and decisions in cardiology.

1.6 Sustainable Development Goals and Social Impact

This research can contribute to SDG 3 (good health and well-being) as shown in Figure 1.5. It improves the accuracy and efficiency of CVD diagnosis through advanced AI techniques to address critical health challenges such as reducing mortality rates and enhancing healthcare outcomes for individuals affected by cardiovascular conditions.

Figure 1.5: Sustainable Development Goals [16]

Furthermore, this research can contribute to SDG 9 (industry, innovation and infrastructure) by promoting innovation in health care through the construction of XAI models. Other significant advocacy was conducted by the WHF and other NGOs together with the WHO in framing the ideas of the national CVD strategies and in translating the global policy into actionable programs. At the core of these endeavors is the participation of stakeholders from communities vulnerable or at risk for CVDs, to ensure that progress in liberating global health contributes also to physical enhancements in the provision and utilization of care at the community and country level. [17]

1.7 Dataset Description

The dataset employed in the research performed within the context of the present proposal includes publicly available PCG recordings downloaded mainly from the PhysioNet/Challenge 2016 database [4]. This dataset is quite famous for designing and building self-driving systems for diagnosing abnormalities in the sound of the heart. It has a total of 4430 PCG recordings which are collected from 1072 subjects from different geographical regions and different recording conditions. This selection of the dataset offers a strong starting point for training the models and assessing the XAI models for the classification of CVDs. The dataset is presented in Table 1.2.

Table 1.2: Statistics of the proposed dataset

Database	**Total Samples**	**Normal**	**Abnormal**
PhysioNet Challenge 2016	3949	3290	659

1.8 Resource Requirement

Python is employed in this research because of its versatility and popularity in machine learning and AI-related projects due to its handy libraries and frameworks. Google Colab, now a cloud-based environment, has enough computational and collaborative facilities that are required for developing and training large numbers of algorithms. It runs Python scripts, has live collaboration features, and is fully compatible with other Google products. In this research, Python along with Google Colab is used for cleaning the data, training the models, and assessing the efficacy of XAI models with Unsegmented PCG signals. Also, there is the use of the tool, Matplotlib/Seaborn in representing data as well as results in graphical presentation forms.

1.9 Organizational Structure of Report

Chapter 1 is the initial chapter and normally provides the reader with the background information on the matter at hand. This study discusses the application domains, project scope, research background, and certain necessities required for this investigation. In the second chapter, a thorough assessment of the literature is done, looking at previous studies and research in the area. The third chapter concentrates on the specifics of the software and model used in the research, as well as how it operates. This includes implementation specifics and algorithms. Chapter 4, which focuses on simulation validation, scientific achievements, and necessary output data is presented using flow charts, figures, and graphs. The research is concluded in Chapter 5 with a summary of the major discoveries and a discussion of any difficulties encountered. It also identifies locations in the field that could benefit from future upgrades. The references are provided at the end by IEEE format.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This section covers thorough information about the ML and DL methods for classifying the heart sound where the data preprocess, feature extraction, and classification techniques are explored in detail. Furthermore, the research focuses on the nature of XAI to analyze unsegmented PCG data with improved results and diagnosis of CVDs.

2.2 Background

The examination of the heart structure accompanied by the classification of probable basic ideas of the pathology is crucial. This portion covers the explanations of normal and abnormal heart sounds, techniques in cardiac examining, auscultation, XAI, and the benefits of applying XAI in the interaction between AI, ML, and DL.

2.2.1 Anatomy of the Heart

One essential organ that is necessary to the circulatory procedure is the human heart. It is in charge of blood circulation throughout the body. The heart's structure consists of several chambers and valves that control blood flow.

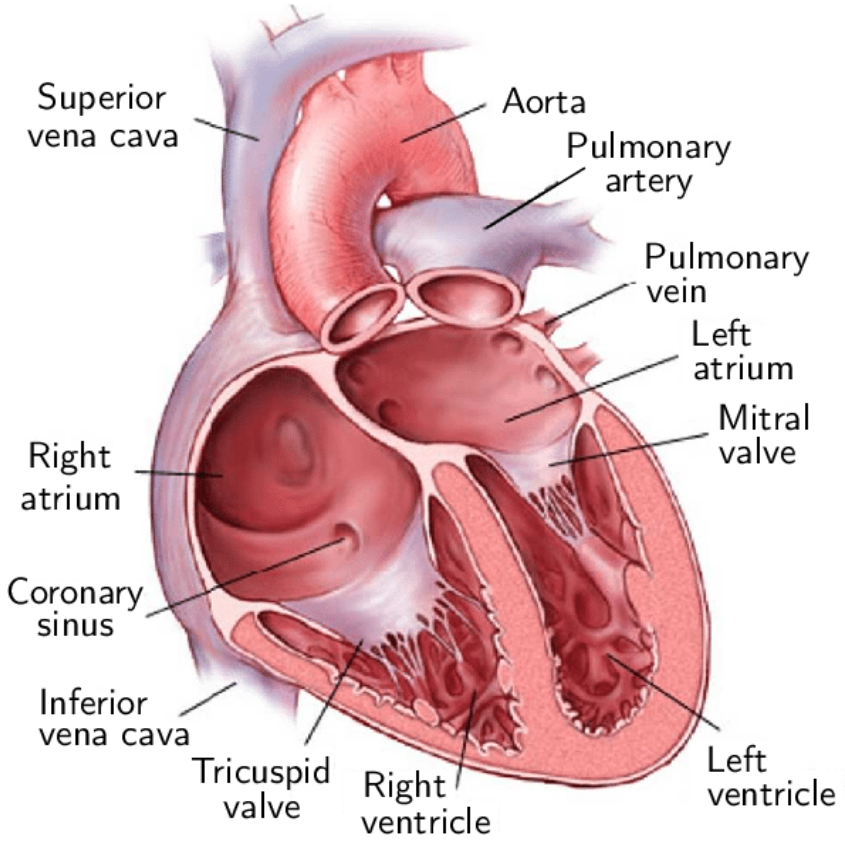

Figure 2.1: The pathway of blood flow through the heart [18]

A specific pathway for blood flow through the heart is depicted in Figure 2.1. First, the deoxygenated blood enters the right atrium. After that, it arrives the right ventricle via the tricuspid valve. Blood is pushed from the right ventricle into the pulmonary arteries via the pulmonic valve. Through pulmonary veins, oxygenated blood is reverted to the heart through the left atrium. After that, it enters the left ventricle through the mitral valve. The oxygen-rich blood is then pumped by the left ventricle via the aortic valve and into the aorta, where it is distributed throughout the body.

2.2.2 Fundamental heart sounds

The auditory sounds and murmurs caused by mechanical events of the heart valves and related vessels are recorded by phonocardiography (PCG) [19]. The auditory sensations of the valvular, muscle, vascular, and blood circulation provide the audible components of heart sound. Physicians can analyze and diagnose various cardiac problems with the use of the PCG signal, which offers crucial clinical information. The initial heart sound, S1, the systolic pause following the sound, the second heart sound, S2, and the diastolic pause following the sound S2 are the four main components of a typical cardiac PCG cycle. Systolic and diastolic interval segments may contain the other additional heart sounds, such as the heart murmurs, the fourth heart sound, and the third heart sound, S3.

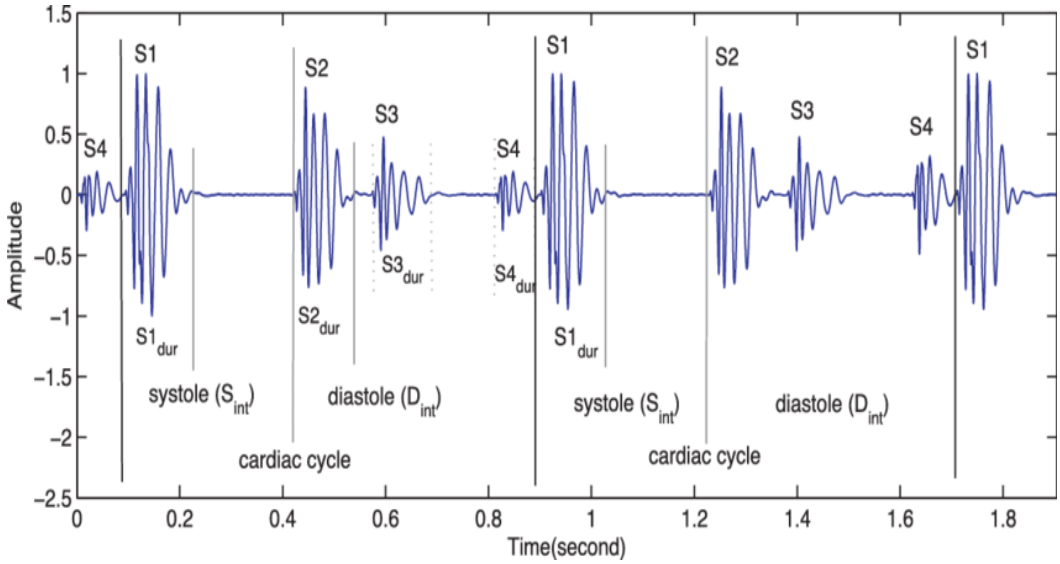

Figure 2.2: Illustrates PCG signal including heart sounds [19]

Figure 2.2 shows a typical heartbeat cycle, highlighting the core heart sounds. S1 is generated as the mitral and tricuspid valves shut and is to be distinguished from the heartbeat which begins at S1. S2 is the end of heart sound this happens when the aortic and pulmonary valves close. S3 occurs in the early resting phase associated with the rapid filling of the ventricle. S4 is conducted in the rest phase occurring towards the end linked to atrial contraction. These sounds assist the doctors in identifying any heart complications by analyzing when they of and how they are produced.

2.2.3 Normal and Abnormal Heart Sounds

It is very relevant to emphasize the significance of heart sounds in assessing the health conditions of the heart. That is because various sound patterns may indicate the state of the organ be it normal or abnormal. These sounds are related to the first and the second stages that is systole and diastole phases. Systole is the stage at which the heart chocks to push the blood out and diastole is the stage at which the heart muscles dilate to allow it to be filled with blood. Said murmurs that exist during these phases yield significant information on the cardiovascular disorder if any exists.

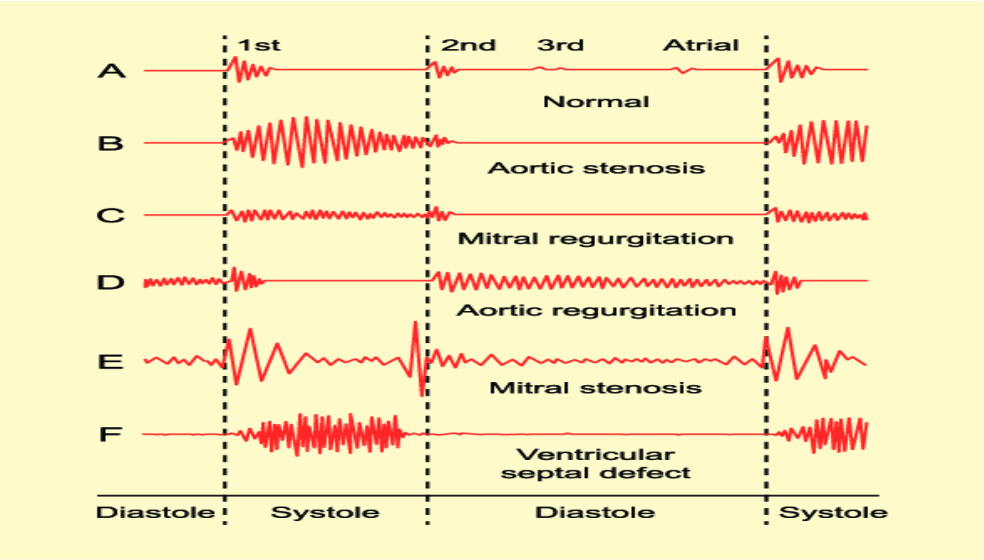

Figure 2.3: Phonocardiogram from Normal and Abnormal heart sounds [20]

Figure 2.3 shows Heart sound patterns of normal rhythmic sounds and abnormal activity due to diseases relating to the heart. The first row represents the normal heart sounds where regular “lub-dub” beats during the cardiac cycle [8]. The second row illustrates the aortic stenosis sound pattern and helps explain an increased, crescendo-decendo systolic murmur . The auscultatory findings for mitral regurgitation include the abnormal heart sound of mitral valve ineffectiveness which is presented in the third row. It is illustrated in the fourth row that aortic regurgitation causes a diastolic decrescendo murmur. The fifth row explains mitral stenosis that traces back to having a diastolic rumbling murmur with

presystolic accentuation. Last but not least, in the sixth row, the learned sound pattern is presented which relates to a ventricular septal defect head a pan systolic murmur. [20]

Table 2.1 summarizes how normal and abnormal heart sounds be able to help identify different heart diseases.

Table 2.1: Summary of Heart Sound Patterns and Associated Heart Diseases

Sound Pattern	Description	Associated Disease
A	Normal heart sound	None
B	High-pitched, mid-systolic	Aortic Stenosis
C	High-pitched, holosystolic	Mitral Regurgitation
D	High-pitched, early diastolic	Aortic Regurgitation
E	Low-pitched, mid-diastolic	Mitral Stenosis
F	Holosystolic, harsh	Ventricular Septal Defect

2.2.4 Cardiac Monitoring Techniques

Cardiac monitoring techniques are crucial for assessing heart health. Electrocardiography (ECG) and Phonocardiography (PCG) are two primary methods used. ECG records electrical activity, showing the heart's rhythm and detecting abnormalities like arrhythmias. In contrast, PCG captures heart sounds, providing information about valve function and abnormal blood flow. PCG is often considered superior to ECG in specific diagnostic scenarios because it directly detects mechanical aspects of heart function that electrical recordings may miss. This makes PCG invaluable for detailed assessment, especially in cases involving murmurs or structural heart issues. Integrating both techniques can provide comprehensive insights into cardiac health, guiding effective diagnosis and treatment strategies [21] . Figure 2.4 shows two graphs that represent normal heart activity.

The top graph is an ECG showing the electrical signals of the heart, while the bottom graph is a PCG showing the sounds the heart makes.

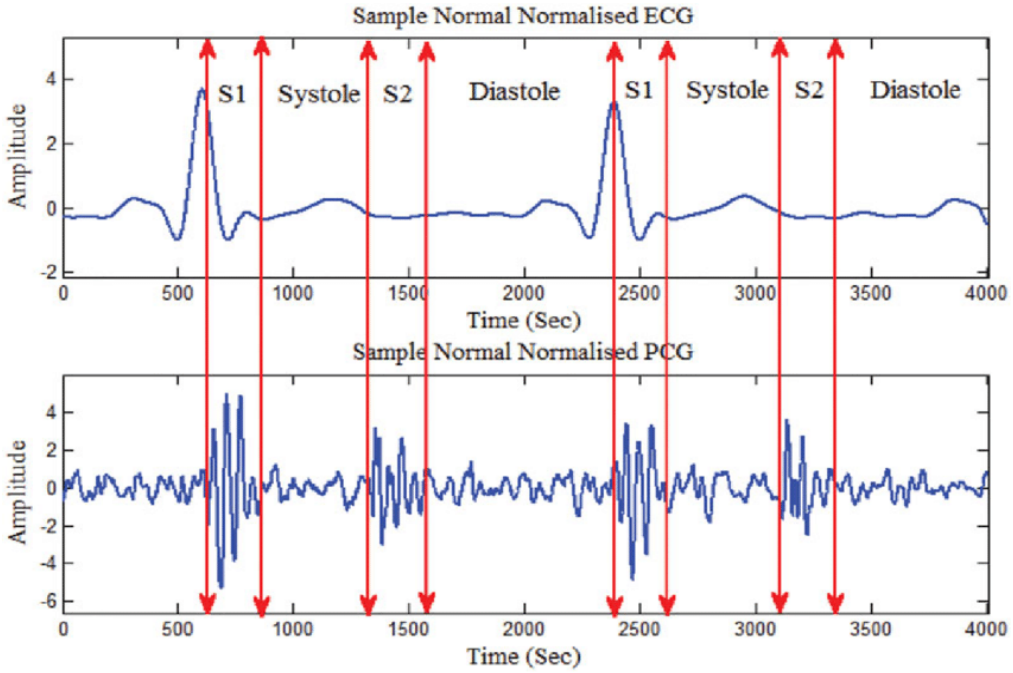

Figure 2.4: Matching two cycles of the ECG with PCG signal [21]

2.2.5 Segmented and Unsegmented PCGs

Segmented PCGs are the actual heart sound recordings that had been preprocessed to isolate the segments of a record corresponding to a single heartbeat or a specific phase of the cardiac cycle. On the other hand, unsegmented PCGs are as recorded, raw heart sound signals without the segmentations. These unsegmented signals consist of all the sounds originating from the heart in a steady stream that does not divide into one or many beats, or segments of the cardiac cycle and all parts, including but not limited to the main ones, S1 and S2, or other noises or murmurs.

2.2.6 Auscultation

Of all clinical applications of stethoscopes, auscultation is the most commonly used technique in distinguishing audible sounds of the heart. But listening to the echoes needs sharp eyes and the help of a cardiologist for proper auscultation [22]. A competent physician when performing auscultation results may be about 80% accurate. Hence, the development of a computer-aided diagnosis (CAD) tool for the assessment of cardiac signals that will assist in a more accurate prediction of cardiac ailments is required. [10]

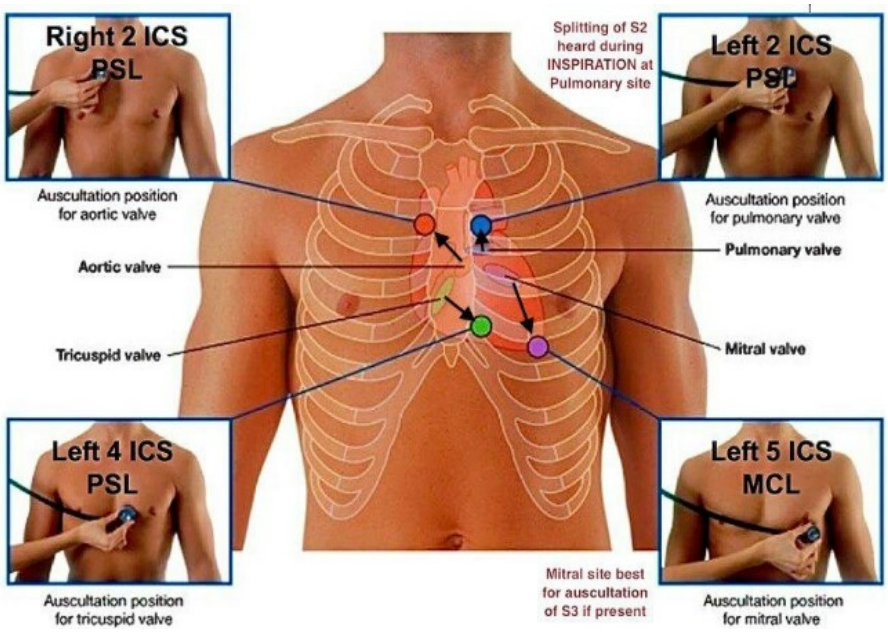

Figure 2.5: Positions for Monitoring PCGs [23]

The illustration in Figure 2.5 [23] gives specific points on the chest where heart sounds are mostly heard when monitoring PCGs through the use of a stethoscope or an electronic sensor. These include the aortic area which is located at the right second intercostal space, the pulmonic area also referred to as the left second intercostal space, a tricuspid area located at the lower left sternal border and the mitral area which is typically the apical part of the heart, usually at left fifth intercostal space. These positions capture sounds for different valves thereby ensuring correct diagnosis of heart diseases.

2.2.7 The Role of Explainable AI in Auscultation

Auscultation, or the act of listening to body sounds, especially the heart and lungs, is a centuries-old diagnostic technique. Auscultation has been one of the major diagnostic tools for heart and breath illnesses. Nowadays, with technological advancements, artificial intelligence (AI) has started playing an increasingly important role in enhancing the diagnostic applications of auscultation. The interpretation of PCGs and some other auscultatory data will be affected by ML and DL, which are among the AI's constituents. However, as these models get complex, there is a need for more transparency and interpretability in them. This necessitates XAI, which provides understandable insights into the decision-making processes.

2.2.8 Benefits of XAI

XAI is useful because it can generate ML models that are transparent, comprehensible, and trustworthy for humans. This value can be utilized in diverse contexts and applications by offering several pluses and advantages.

Table 2.2: XAI Benefits

Benefit	Description
Improved Decision-Making	Offers insights to enhance decision-making
Increased Trust and Acceptance	Builds trust by making AI models clear and understandable
Reduced Risks and Liabilities	Addresses regulatory and ethical concerns, reducing risks.

The main advantages of XAI are explained in the table 2.2. It demonstrates how XAI lowers risks by addressing ethical and legal issues, enhances decision-making by providing insights, and fosters trust via transparency.

2.2.9 Principles of Explainable Artificial Intelligence

A collection of rules and suggestions known as XAI principles can be applied to the creation and application of transparent and understandable ML models. These guidelines can guarantee that XAI is applied responsibly and ethically while also offering insightful information and advantages across a range of fields and applications.

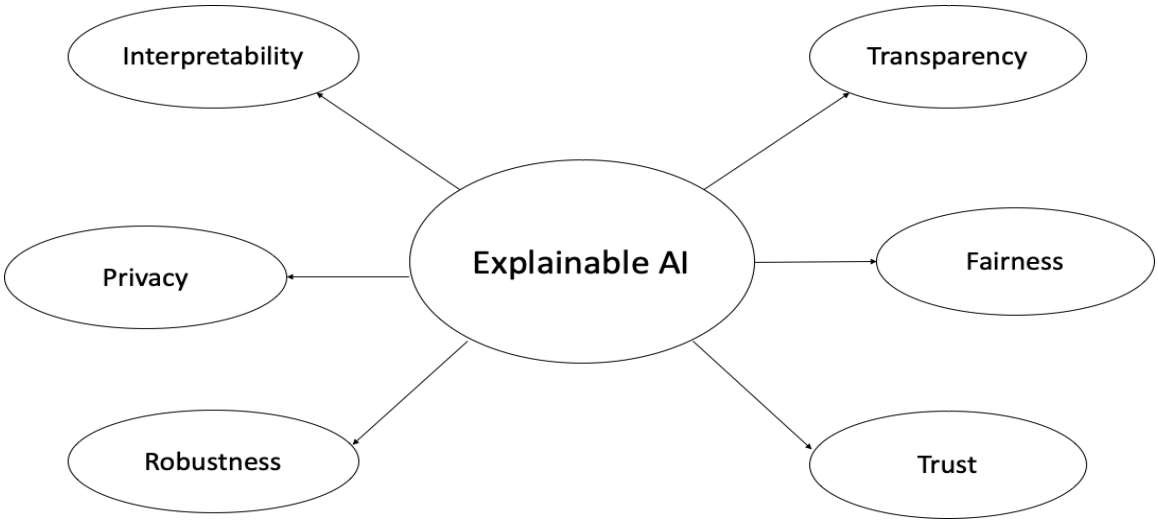

Figure 2.6: Key Principles of Explainable AI (XAI) [24]

Building trust in AI systems requires understanding the fundamental XAI principles, which are illustrated in Figure 2.6. Encouraging clarity and openness plus making sure stakeholders can comprehend the model's decision-making processes are all part of transparency. To prevent discrimination based on race, religion, gender, disability, or

ethnicity, fairness focuses on making sure that model decisions are impartial and equal. Assessing and verifying the degree of trust that human users have in the AI system and promoting reliance on its results are two aspects of trust. To retain consistent and dependable performance even in the face of uncertainty or unforeseen circumstances, a model must be robust to changes in input data or parameter values. Sensitive user data is protected against data breaches and misuse according to privacy assurances. Interpretability is the process of giving people an easy-to-understand and valid explanation for the predictions and results of the models, hence facilitating AI-driven decision-making. When taken as a whole, these guidelines improve the AI system's dependability, equity, and transparency, especially in vital applications like healthcare. [24]

2.2.10 XAI Techniques

It is crucial to classify the different XAI strategies according to how they provide interpretability to comprehend them. The various XAI strategies are broken down into transparent methods and post-hoc methods in Figure 2.7.

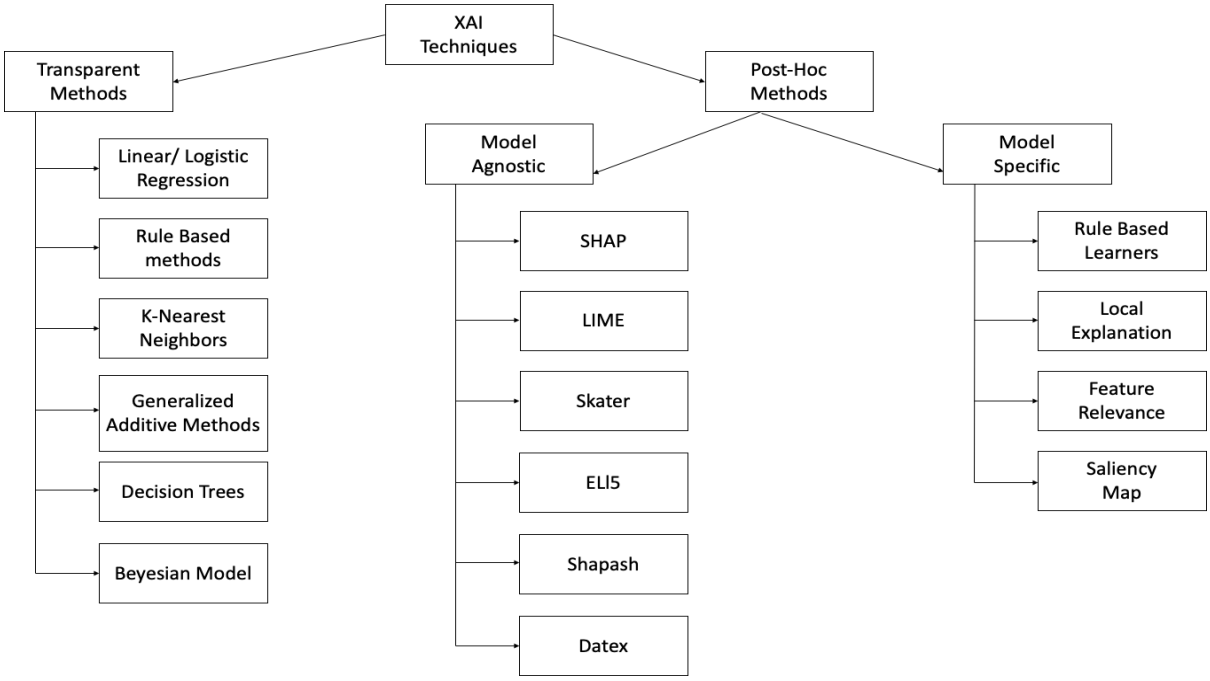

Figure 2.7: Explainable AI Techniques [14]

The two primary categories for XAI techniques are Transparent Methods and Post-Hoc Methods, as shown in Figure 2.7. Techniques that are naturally interpretable, such as decision trees, generalized additive models, k-nearest neighbors, rule-based approaches, Bayesian models, and linear/logistic regression, are examples of transparent methods. After the model is constructed, Post-Hoc methods which are separated into model-specific and model-agnostic approaches offer explanations. Model-specific techniques like rule-based learners, local explanations, feature relevance, and saliency maps are designed specifically for a given model to provide insights into its inner workings, whereas model-agnostic techniques like SHAP, LIME, Skater, ELI5, Shapash, and Dalex can be applied to any kind of model.

2.2.11 The Integration of AI, ML, DL, and XAI

It is essential to understand how concepts interrelate and overlap within the broader field of AI. The below figure visually represents these relationships.

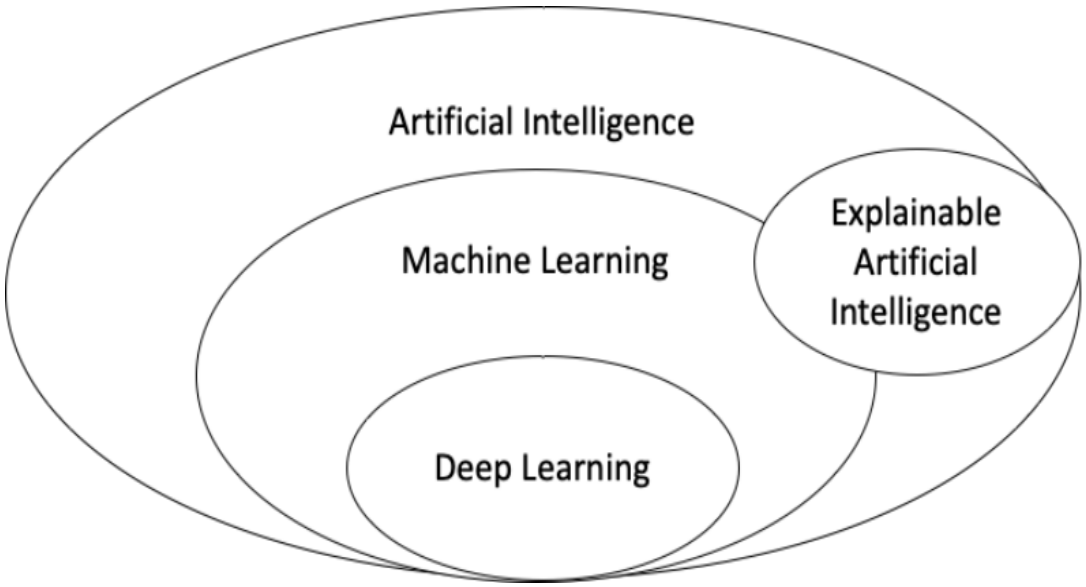

Figure 2.8: Relationship Between AI, ML, DL, and XAI [25]

The ML functions as a specialized subset that creates algorithms for computers to learn from data, Figure 2.8 depicts the hierarchical and overlapping links between major concepts in AI, demonstrating how AI spans the broad subject of developing intelligent machines. DL is a subfield of ML that uses neural networks to process intricate pattern recognition problems. Additionally, it draws attention to XAI, which combines elements of ML and DL. The interconnectedness of AI, ML, DL, and XAI is emphasized by this tiered paradigm, which also highlights the significance of explainability in sophisticated AI models.

2.3 Literature Review

This portion reviews the literature on the methodologies used in the identification of cardiovascular diseases, with subtopics on machine learning, deep learning, and a combination of both. The real use of XAI techniques as they relate to CVDs is also examined, focusing on the significant role that XAI plays in enhancing model explainability and advancing clinical decision processes.

2.3.1 Machine Learning Techniques

Machine learning techniques were used in the study [10] for the classification of cardiac sounds to identify anomalies. The study highlights the need for early identification of heart disease, a major world health concern. The publicly available dataset Physionet/ Challenge 2016 was utilized by the researchers for determining heart abnormalities. Unlike conventional techniques, the method analyzes the whole raw PCG signal for feature extraction, omitting the segmentation stage. The main strategies used are classification and preprocessing by using unsegmented PCGs. This work primarily concentrates on removing the segmentation process for increased efficiency. During preprocessing, the raw PCG signal is directly analyzed using a variety of methods, including Power Spectral Density (PSD) to show power distribution across various frequencies. Mel-Frequency Cepstral Coefficients (MFCC) focusing on human perceivable frequencies, and Homomorphic Envelogram to

comprehend amplitude variations over time. The Ensemble Boosting method which is also known as AdaBoost is used for classification, many weak learners are combined into a stronger classifier for improved heart sound abnormality detection. This method demonstrates its promise for early disease detection and improved patient care with high accuracy and sensitivity for abnormal identification. However, trustworthiness was not established for these classifiers by the researchers.

In the article [4] applications of machine learning are examined to label heart sounds. They are divided into two most important stages of the exploration which are classification and data preprocessing. During the preprocessing phase, all audio files are resampled to a standard frequency of 2 kHz (2000 Hz) to sustain equality between recordings. An automated process is involved for labeling and estimating quality, monitored by a manual accuracy check. Authors used Binary Logistic Regression which is the main method for classification. When there are two possible consequences this statistical method works well. To define normal and abnormal heart sound classes it examines features recovered from segmented heart sound recordings. Other classification strategies were discussed by the authors for further investigation, such as clustering techniques for organizing related recordings based on shared characteristics, Support Vector Machines (SVMs) for determining the best separation boundaries between classes, and Hidden Markov Models (HMMs) for detecting sequential patterns within the sounds.

Researchers investigate the ML algorithms in the study [26] for the identification of cardiac illness by using the Physionet dataset. The data set contains 2435 annotated heart sounds that are classified as normal or pathological by combining recordings from multiple sources. The vital part of research is preprocessing which is done on the heart sound recordings. The following steps were involved in the preprocessing, which are segmentation, fragmentation, peak identification, and noise reduction. These techniques eventually increase the precision of illness detection by cleaning and preparing the data for feature extraction. Support vector machines (SVM), Naive Bayes (NB), and Decision Trees (DT) were investigated by researchers. They test diverse feature sets, including short-term, long-term, and combination variants, that were taken from the preprocessed data. Interestingly, the outcomes display the classification correctness of all tested algorithms is significantly increased.

The focal issue of the literature review in the study [27] is the classification of heart sound signals with the use of the dataset Physionet. It is a broad dataset with exceeding than 2000 records of healthy people and heart disease patients. To access the recommended classification approach a subset of 409 heart sound recordings were chosen for the identification. The dataset is split into training and testing sets to provide a strong groundwork for evaluating the algorithms.

Features were extracted with the help of wavelet scattering transform. In addition to recovering high-frequency data that would have been lost during low-pass filtering, this technique produces stable signal features. The produced scattering coefficients, are especially useful for the challenging task of classifying heart sounds because they offer elastic deformation stability and local translation invariance. The integrity of the signal characteristics, which are essential for precise classification, is preserved by using this method.

To increase the categorization correctness and computational efficiency a ML method Twin Support Vector Machine (TWSVM) is used for classification. To address the high dimensionality of the feature vectors that are developed from the wavelet scattering transform this study uses Multidimensional Scaling (MDS) for dimensionality reduction. The combination of MDS and TWSVM is contrasted with other combinations, such as MDS and SVM, PCA and TWSVM, and PCA and SVM. According to the results, a combination of MDS and TWSVM achieves superior than the other techniques in relations of competitive running time and classification accuracy. This explains the worth of the recommended approach, which might be improved yet by growing the dataset and fine-tuning TWSVM settings in subsequent studies.

The study [28] presents a novel approach to heart sound categorization that does away with segmentation with PCGs. The study uses wavelet decomposition, which significantly increases the classification model's accuracy to extract features from unsegmented PCG recordings. The research attains remarkable performance metrics utilizing the K-Nearest Neighbors (KNN) classifier. Strong classification performance while cutting down on complexity and time for early detection and treatment of cardiovascular illnesses is delivered by this research.

To reduce high and low-frequency noise from the PCG signals a fourth-order Butterworth bandpass filter with cutoff frequencies of 25 Hz and 400 Hz is used for preprocessing. Likewise, a spike removal technique is applied, which entails partitioning the PCG recording into 500 ms intervals, determining the highest absolute amplitudes, and eliminating spikes according to predetermined standards. The dataset used in the study is the Physionet 2016 challenge, which is categorized into several subsets and designated as dataset 'A' through dataset 'F', represented. There are 3240 recordings in all in this dataset, which represents actual situations that are experienced during auscultation. Using data segregation for classification the artificial neural network (ANN) model is tested with 15% of the data, validated with 15%, and trained with 70% of the data.

The authors also intimated K-nearest neighbors (KNN) classifier to categorize heart sounds which is the foremost purpose of the effort. Using extracted features, offering high categorization accuracy rates the KNN method is used for unsegmented PCG recordings. The study also comprises the use of unsegmented feature-based decision tree classifier procedures, which produced results with 77% accuracy and sensitivity. KNN performed better in terms of accuracy and sensitivity as matched to other classification techniques such as ANN. It explains how well it can categorize unsegmented heart sound recordings to pinpoint cardiovascular diseases early on.

Deep learning tactics are not covered in the study. The authors don't use Explainable Artificial Intelligence (XAI) approaches. Enhancing the transparency and interpretability of the classification process could be achieved by incorporating XAI techniques like model-agnostic interpretability approaches or feature importance analysis.

2.3.2 Deep Learning Techniques

In a study [1] authors used the Unidentified PCG recording dataset. Insufficient healthcare resources in low and middle-income nations are tackled. It also intimated deep learning techniques for automated heart sound categorization. To prevent the human feature engineering and segmentation that is frequently employed in conventional methods, the

advised approach pursues to automate feature extraction and classification. Preprocessing entails segmenting the raw signal into 6-second epochs for analysis and down sampling it to lower computational demands. Furthermore, a Savitzky-Golay filter is employed to cut high-frequency interference extant in the stream.

Distinctive natures of deep neural networks e.g. one-dimensional convolutional neural network (1D-CNN) and a five-layered feed-forward neural network (F-NN) are paralleled. With an accuracy of 85.65%, the F-NN model overtook the others and indicated that deep learning is feasible for real-time heart sound classification without the need for segmentation or manual feature engineering. This technique has the prospective to improve patient care in situations with limited resources by enabling early illness identification.

The authors categorized heart sounds in the study [29], and spot anomalies by using deep learning techniques. It comprehends heart illness as a global health concern and emphasizes the prominence of observing heart function for preventative interventions. The publicly available dataset PhysioNet/Challenge 2016 is used for the classification of heart abnormalities. Researchers used a multi-step method called preprocessing and classification. The Discrete Wavelet Transform (DWT) approach is used for multi-resolution analysis of the raw PCG signals. In addition to lowering noise, this efficiently squeezes the data.

Once removing noise the signal is then split into individual segments, each of which represents a diverse component, Segmentation is accomplished by examining the signal's energy distribution and zero-crossing locations. Finally, Mel-scaled spectrograms and Mel-frequency cepstral coefficients (MFCC) are used to extract relevant features from the segmented signal. By converting the signal from time domain into a visual depiction of the frequency content, these approaches effectively capture features that are pertinent to human hearing. Five layers feed-forward Deep Neural Network (DNN) a deep learning model is used in the study for the classification of Phonocardiograms. This joint approach of deep learning and signal processing methods demonstrates encouraging potential for analyzing and classifying PCG signals. Such improvements might contribute to the timely detection of heart abnormalities, paving the way for upgraded patient care.

In the context of telehealth, a deep convolutional autoencoder (DCA) to deal with phonocardiography is investigated in the work [30]. The Dalian University of Technology heart sounds database (DLUTHSDB), which includes recordings from both healthy individuals and patients, is the open-source dataset that the study uses from Physionet. This study also suggests a method for compressing cardiac coronagraph (PCG) information via Deep Canonical Correlation Analysis (DCA). The S1 and S2 heart sound segments are the central focus of the segmentation, overlapping windows, and normalization techniques used by the system to preprocess the PCG signals. The DCA, the system's central component, compresses the signal before sending it to distant medical specialists. By gradually lowering the dimensionality of the signal representation in an encoder and then reconstructing the signal in a decoder, the DCA architecture accomplishes compression.

The authors also conclude that a segment length of three seconds provides a good deal between signal quality and compression ratio. They were able to attain a 32-compression ratio with less than 5% PRD in the signal. They also investigate the system's resilience to transmission failures and whether it can be used on devices with less processing capability. The outcomes demonstrate that when compared to conventional compression methods, the DCA method is more noise-resistant.

In an article [31], researchers research automated heart sound classification using deep learning to improve cardiovascular disease (CVD) screening. Cardiovascular diseases (CVDs) are a commanding foundation of death internationally. Although traditional cardiac sound analysis has its value, it is subjective and demands a great deal of knowledge. The latest studies use automated screening methods for the classification of CVDs either rely on complex models or demonstrate inefficiencies. The study used two publicly available datasets for the classification of PCGs. The PhysioNet 2016 and PASCAL 2011 to determine CVDs. However, this is problematic since there are differences in the methods used to collect and examine the data, as well as noise in the recordings. In order to overcome these issues, the writers use the short-time Fourier transform (STFT) to change the PCG signals into spectrograms.

On the PhysioNet dataset, recommended CNN model performs admirably, outperforming previous techniques in accuracy and using less processing power. When the

datasets are combined, performance remains good. Transfer learning allows the model to attain good precision in heart sound classification even on the noisier PASCAL dataset. This study shows the possibility of transfer learning in conjunction with a custom CNN model that is less refined for the perseverance of CVD screening utilizing PCG data.

An innovative deep learning method for automatic identification of phonocardiogram (PCG) signals is being studied in research [32]. To diagnose cardiac valve dysfunction with PCG analysis, these critical sounds must be identified early and accurately. Conventional techniques for FHS detection can be ineffective and time-consuming.

The paper also suggests a time-frequency domain (TFD) deep neural network (DNN) method to challenge this problem. The PCG signals are altered into a time-frequency representation using the MGWST (Modified Gaussian window-based Stockwell Transform), which offers a more detailed understanding of the signal's properties. The study practices a multi-step method to obtain vital attributes. Primary, the segmented heart sound issues are evaluated using the Shannon-Teager-Kaiser energy (STKE). These are then further refined through the application of smoothing and thresholding procedures. Ultimately, the segmented components are used to extract time-frequency Shannon entropy (TFDSE) characteristics.

The occurrence or lack of FHS components in the PCG signal is then spontaneously categorized using a DNN architecture grounded on stacked autoencoders (SAEs). With noteworthy outcomes, the efficacy of this strategy is assessed on two publically accessible databases Database 1 is the Michigan Heart Sound and Murmur Database, and Database 2 is PhysioNet Computing in Cardiology Challenge 2016. The authors suggested that the deep learning strategy works better for FHS detection in PCG signals than current techniques. It also delivers an actual defined and efficient approach for PCG recording analysis, which allows the potential for real applications in the early association of cardiac disease.

2.3.3 Hybrid Techniques

In direction to assist in the initial finding of heart disorders, the article [33] specifies a different approach for the categorization of heart sounds. The three primary chunks of the recommended method are the creation of the image, feature extraction using CNN models that have previously been trained (Alex Net, VGG16, and VGG19), and feature classification using an SVM classifier. Spectrogram pictures are produced for extracting features, from input heart sound waveforms. The frequency content of the signal is represented over time by spectrogram pictures, which are produced by means of Short Time Fourier Transform (STFT). The research leads that deep learning methods have the potential to recover cardiac disease recognition systems, as perceived by the improved performance observed in comparison to current methods.

In the article [34] to categorize heart sounds and subsequently spot malfunctions, insufficient different methods are applied. The Amalgamation of Machine Learning and Signal Processing is used. Numerous methods have been proposed and advanced by researchers to develop ways of heart sound classification, however in general they contain the preprocessing, segmentation, feature extraction, and classification stages. Some of these techniques have been examined for segmentation and classification, such as deep learning algorithms, Hidden Markov Models (HMMs), and Hidden Semi-Markov Models (HSMMs). The dataset is preprocessed by operating different preprocessing methods such as resampling, normalization, and filtering, alongside machine learning methods such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees (DT) used for classification.

Heart-related illnesses are the principal killers worldwide, and the importance of spotting, diagnosing, and treating them is well caught in the article [35]. Thus, there is a note that new approaches to automated methods of heart sound diagnosis imply segmentation of Phonocardiogram (PCG) signals that increase examination complexity and computational limitations. The paper is going to replace segmentation and assess the advantages and disadvantages of the proposed method for establishing short, unsegmented 5-second PCG recording data. It also focuses on a new method for heart sound classification applying a pre-

trained convolutional neural network (CNN) model to the PhysioNet2016 challenge. Before feeding, the short PCG recordings, authors apply a preprocessing step called continuous wavelet transform (CWT), which creates 2D Scalo gram images out of the signals which are further employed in training and testing the CNN model. Such a method is advised to have rarer difficulties in relation of segmentation complexity but is also quantified to show good results linked to existing methods in the framework of heart sound classification.

In a study [36] researchers focus on the work to design machine learning algorithms to particularly advance the examination of unsegmented PCG signals for edge computing solutions in wearable healthcare devices. The study makes more use of computationally less complex classifiers such as SVMs, feed-forward NNs, and k-NNs that work well and are not overly complex. In this context, it emphasizes that the models that hinge on the segmentation algorithms or the classes of methods pointedly exhaust the resources.

It also used the PhysioNet 2016 dataset, employing preprocessing techniques such as spike removal, noise reduction, normalization, and segmentation into 5-second frames, and utilizing classifiers like SVMs, NNs, and k-NN. A notable highlight of the study is its focus on classifiers that are suitable for resource-limited platforms, particularly underscoring NNs. The feasibility of implementing these trained classifiers on platforms like Raspberry Pi 4 or NVIDIA Jetson Nano is discussed, showcasing potential advancements in real-time monitoring and diagnostics of heart conditions through wearable technology. Overall, the findings propose promising implications for the swift utilization and scalability of diagnostic devices in the future healthcare landscape.

The article [37] focuses on applying Deep Learning with Traditional ML methods to advance the results of Phonocardiogram (PCG) signal categorization. Thus, through the use of DL in extracting the features from unstructured data and ML in avoiding overfitting of the proposed deep hybrid models, the concept proved potential in identifying heart diseases at the initial stages. It expresses an architectural model including convolution layers, dropout layers, and fully connected layers, and thereafter the prediction classes are managed by Machine Learning classifiers. Using performance evaluation, it is possible to achieve better accuracy than using Deep Learning or Machine Learning only. The deep hybrid models possess computational gains and less time complexity and they do not require feature

engineering hence deep hybrid models could be very useful for the early detection of cardiovascular anomalies. However, those comprise plentiful limitations and wrap up important aspects of Explainable AI (XAI) that are not included in the study, such as the features by which the models make decisions.

The dataset used in the work is the PhysioNet 2016, which consists of a large number of records of PCG, and this allows for the development of a solid framework for the advancement of the proposed models. The implementation of this dataset gives credit to the reliability of the results that are derived from the quality and the variety of datasets that will be used to build the models that will be generic to various patients. Due to the detailed annotation and validation of the proposed PhysioNet 2016 dataset, this dataset can be recommended for use in scientific work, due to this it is possible to check the effectiveness of the newly developed methodologies.

The article's findings validate that the collaboration of DL and Traditional ML methods can go an extensive way in producing monumental progress. The hybrid approach described in the study also helps to improve the classification and minimizes some of the shortcomings that exist when using each method independently. While on one hand Deep Learning does not require much preprocessing of raw data to extract features for the model, a problem that Traditional Machine Learning effectively solves is that of overfitting, which is prevalent in large models.

However, it should be observed that there is no application of XAI techniques which is a limitation given the significant outcomes of the study. The concept of Explainable AI becomes vital in such areas as healthcare since the AI model must be easily understood by healthcare providers to gain their confidence and ensure that they embrace the technology. Without XAI, the deep hybrid models have the characteristics of a 'black box,' which provides minimal information regarding decision-making. This lack of transparency overly impacts the application of these models in real practice because clinicians may not be willing to work on a tool whose results they do not understand.

2.3.4 XAI Techniques for CVDs

The audio databases used in the article [15], PhysioNet is a versatile set of records that captures the heart sound in clinical-nonclinical situations. However, this distribution is skewed, which means, there must be a larger sample of the normal heart sound than the abnormal ones. About this, and to improve the data for analysis, the authors used Mel-Frequency Cepstral Coefficients (MFCCs) for feature extraction. These features express the spectral properties of the heart noises. For classification, the inclusion of a Classification Network using a CNN connected to an MLP is used in the study. This model yielded good results in discriminating between normal and abnormal sounds of the human heart.

Moreover, the research does not only entail just classification but includes the use of Explainable AI (XAI) methods. SHAP (Shapley Additive explanations) and Occlusion maps are used to explain how the model views the data and how it comes to those conclusions. It is especially important as far as the medical field is concerned, where people do trust the AI models.

In [38] the authors try to assess the deep learning models in detecting abnormal sounds from the PhysioNet database. Another important aspect that the study successfully grips is the class imbalance in the data set meanwhile windowing is used to get signal segments with one-second intervals but with overlap and then balancing is also done. Three deep learning frameworks are discussed composed of different parts for feature extraction and classification. One model employs a segmentation model that is pre-trained with a CNN together with the use of a CNN encoder and an MLP classifier. Another model has connections between the CNN encoder and MLP network but adds features obtained from the segmentation model as another channel of input. Last, a CNN-MLP network without segmentation inputs is created separately from the previous networks.

This type of research focuses on the necessity of explainability in medical uses of AI. Using Shapley values, the analysis offers information on the influence of specific characteristics of the signal within the heart sounds on the decision made by the model regarding the existence of abnormalities. This tends to facilitate an understanding of how the

model's decisions are arrived at. Also, occlusion maps are employed to depict areas of the heart sound signal that contain the most critical information for predictions. Last but not least, The performance of all the models is assessed with 10-fold cross-validation.

The growing incidence of CVDs and its effects on public health have made the classification of these disorders an important field of study in topical years. Conventional methods for diagnosing CVDs, such as ECGs, concentrate on capturing the heart's electrical activity. Nevertheless, there are situations where these techniques are unable to identify minute alterations in cardiac function and mechanical anomalies. This has increased interest in PCG, a complementary method that can offer further insights into heart health by recording the noises made by the heart. [21]

Advancements in AI have further propelled the medical field, with ML techniques being applied to classify and diagnose CVDs. Though, one of the main challenges with AI-based methods is the lack of transparency, often described as the "black box" problem, where the decision-making process is not easily understood. So there is a pressing need for XAI models in the medical domain. To address this issue, researchers have turned to XAI, which aims to make AI models more interpretable and their decisions more transparent. [15]

A stethoscopy or heart auscultation is also individual of the widely used approaches for the diagnosis of heart diseases, which is very safe and inexpensive however, the effectiveness of this approach is predetermined by the individual characteristics of people's hearing, the quality of stethoscopes, and physicians experience. In some pathological situations, the evaluation of essential heart sounds is masked by high-frequency murmur eventually making diagnosis difficult. Regarding this, physicians turn to computer-aided heart sound analysis (CAHSA) systems which assist in interpreting, auditioning, and visualizing intricate heart sound signals. [19]

However, in the current state, there are some main issues related to the classification of CVDs using unsegmented PCG signals. Discontinuity of heart sounds, inter-subject differences and the formation of various physical noises from respiration and neighboring environmental sounds present an immense impact on the PCG analysis. Furthermore, the

classical methods of the segmentation of ECG signals might be significantly worse, for instance, in cases when children, especially newborns, are involved or when the background noise is high. To address these challenges, this research revolves around creating ML algorithms for the categorization of CVDs from unsegmented PCG signals and creating an XAI for these algorithms.

2.4 Summary

The literature review is summarized in Table 2.3. The condition that is mainly covered is the ML and DL methods for heart sound classification. Sophisticated areas for further research are data preprocessing, features and extracting them, and classification. As for the methods, binary logistic regression, SVM, Random forest KNN, CNN, and DNN are used. These techniques are applied to datasets like PhysioNet/Challenge 2016 to enhance the classifier's performance and distinguish between healthy and unhealthy heart sounds. From the analysis of the results obtained by employing the combination of the ML, the DL, and the XAI methods, future achievements regarding the diagnosis of heart diseases and the improvement of the treatment of patients, can be expected.

Table 2.3: Summary of Literature Review

Ref	Preprocessing	Dataset	Classifier		
			ML	DL	XAI
[1]	Down-sampling, Segmentation, Savitzky-Golay filter	PCG recordings	✖	✔	✖
[10]	Homomorphic Envelogram, Mel-Frequency Cepstral Coefficients (MFCC), Power Spectral Density (PSD)	2016 PhysioNet/CinC	✔	✖	✖

[15]	Mel-Frequency Cepstral Coefficients (MFCCs)	PhysioNet Database	✖	✔	✔
[26]	Noise Reduction Peak Detection Segmentation Fragmentation	2016 PhysioNet/CinC	✔	✖	✖
[27]	Wavelet Scattering Transform Low Pass Filter	2016 PhysioNet/CinC	✔	✖	✖
[28]	Filtering, Spike Removal	2016 PhysioNet/CinC	✔	✖	✖
[29]	Discrete Wavelet Transform (DWT), Segmentation, Mel-scaled power spectrogram, Mel frequency cepstral coefficients (MFCC)	2016 PhysioNet/CinC	✖	✔	✖
[30]	Normalization, Segmentation, Overlapping segmentation, Feature selection	PhysioNet Dalian University of Technology heart sounds database (DLUTHSDB)	✖	✔	✖
[31]	Short-time Fourier transform (STFT)	PASCAL, 2016 PhysioNet/CinC	✖	✔	✖
[32]	Modified Gaussian Window-based Stockwell Transform (MGWST)	2016 PhysioNet/CinC , Michigan heart sound and murmur database	✖	✔	✖
[33]	Spectrogram generation, Short Time Fourier Transform (STFT)	Classifying Heart Sounds Challenge (CHSC)	✔	✔	✖

[34]	Resampling, normalization, and elliptic filter, Bandpass filter	2016 PhysioNet/CinC	✔	✔	✖
[35]	Filtering Spike Removal Continuous Wavelet Transform (CWT)	2016 PhysioNet/CinC	✔	✔	✖
[36]	Spike removal, Noise reduction, Normalization	2016 PhysioNet/CinC	✔	✔	✖
[37]	Frame Division, Fourier Transform, Power Spectrum Calculation and Mel-Scale Filter Banks, Logarithmic Energy Summation, Mel-Frequency Cepstral Coefficients (MFCC)	2016 PhysioNet/CinC	✔	✔	✖
[38]	Windowing Data Balancing	PhysioNet Database	✖	✔	✔
[39]	Resampling, Data Labeling and Correction	2016 PhysioNet/CinC	✔	✖	✖

It is observed in the above studies that substantial amounts of heart sound data are generated, posing a significant challenge for efficient analysis and classification. The foremost inspiration of the research is to inspect the CVDs using XAI based on unsegmented PCGs.

CHAPTER 3

METHODOLOGY

3.1 Overview

This section offers a detailed approach to heart sound categorization, emphasizing the use of ML and XAI approaches on unsegmented PCG data. It describes the techniques used for feature extraction, classification algorithms, and data preprocessing. Furthermore, the description of integrating XAI techniques to improve model interpretability and transparency offers a thorough foundation for the analysis and validation of the suggested strategy that follows.

3.2 Proposed Methodology

Figure 3.1 depicts the recommended methodology. It involves the subsequent subtasks such as the publicly accessible PhysioNet 2016 dataset used and arranging it in a format that is suitable for additional processing constitutes the first stage. The dataset must next undergo the required preprocessing and noise reduction. Preprocessing is followed by feature extraction, which addresses class imbalance. The Random Forest classifier was initialized with 100 estimators and trained alongside support vector machines (SVM) and k-

nearest neighbors (KNN) on the scaled training data. The performance of the classifiers was evaluated on the test set, using a confusion matrix to calculate key metrics such as accuracy, precision, recall, specificity, and F1 score, providing a comprehensive assessment of the model's effectiveness. Additionally, the transparency of the categorization process is enhanced by the usage of XAI techniques. This comprehensive approach ensures both good model performance and interpretability.

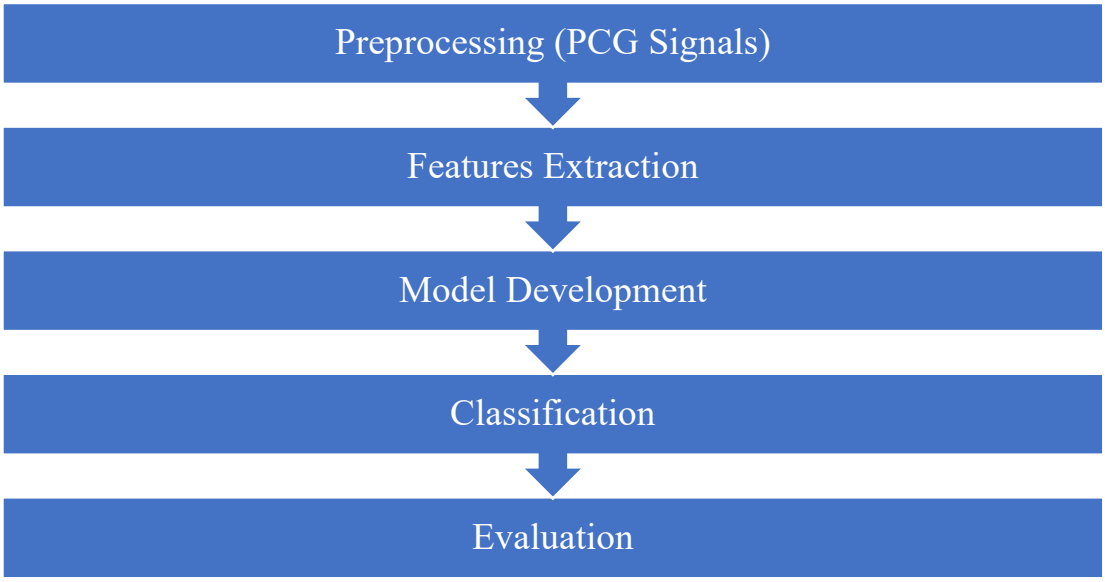

Figure 3.1: Methodology

3.3 Preprocessing of PCG Signals

This is a vital step in the consideration of PCG signals as it prepares the raw data for more accurate and effective model development. The preprocessing of PCG signals is essential for ensuring accurate and reliable analysis. This process in the proposed research includes two main steps: signal denoising and handling class imbalance.

3.3.1 Signal Denoising Using Butterworth Filter

This study makes use of a Butterworth bandpass filter to improve physiological signal quality by lowering noise and keeping required signal components inside a given frequency range. Previous studies [31] show that primary heart sounds and murmurs lie in the frequency range of 20 to 400 Hz. As seen in Figure 3.2, a fourth-order Butterworth bandpass filter was applied with cut-off frequencies ranging from 20 to 400 Hz. This method efficiently suppresses noise artifacts while maintaining the retention of pertinent signal components in CVD classification, particularly in PCG signal analysis [40]. By restricting the passband to 20 Hz to 400 Hz, the study concentrated on the crucial frequency components associated with cardiovascular signals, such as heart sounds and murmurs.

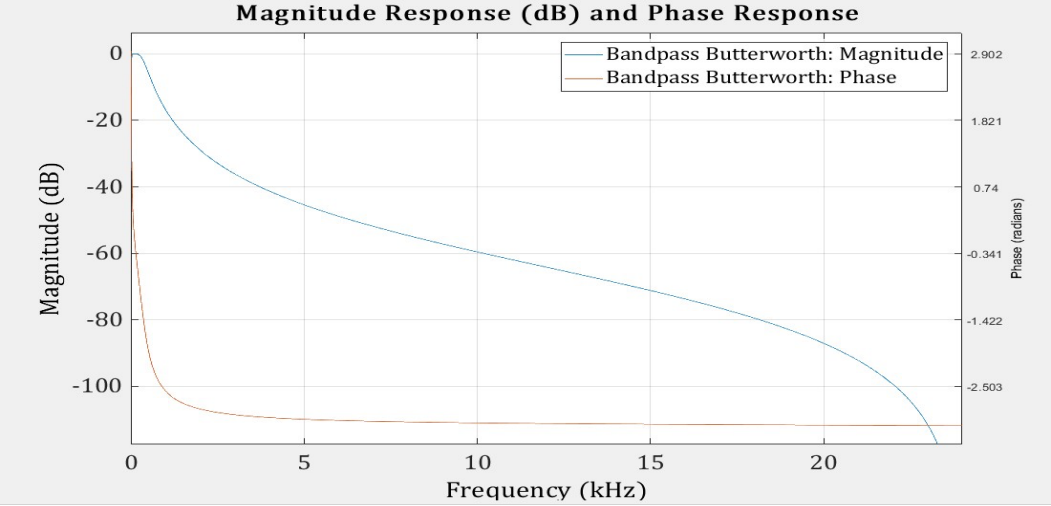

Figure 3.2: Magnitude and Phase Response of the Butterworth filter [31]

3.3.2 Handling Class Imbalance

One of the critical challenges in the proposed dataset is class imbalance, where definite classes are underrepresented and paralleled to others. Initially, the dataset consisted of 659 instances labeled as patients and 3,290 instances labeled as healthy, highlighting a

significant class imbalance. This disparity could potentially bias the model towards the majority class during training and evaluation.

To address this problem and consequently, fairly measure the performance of the constructed model, the identified research employed the Synthetic Minority Over-sampling Technique (SMOTE). Using the minority class (label 1), SMOTE was used to create synthetic samples by interpolating between instances, and this led to there being an equal amount of instances for both classes after SMOTE. To achieve this, after applying SMOTE the number of instances for both patients and healthy people was balanced to be 3290.

Regarding SMOTE, this rebalancing strategy is significant as it decreases the model's likelihood of being imbalanced in terms of the majority class. The incorporation of SMOTE into the preprocessing pipe has shown an appreciation towards dimensioning the work on the handling of class imbalance problems and enhancing the soundness of the model in its recommendation.

Thus, the proposed research integrates SMOTE with denoising methods in the preprocessing stage which will pave the way for the rest of the analysis and model formulation. This approach not only addresses the class imbalance issues but also leads to more accurate and robust solutions from the predictive models in healthcare data mining.

3.4 Feature Extraction

In this step, an overall 13 features were extracted as mentioned in table no. 3.1 below. These features reflect different kinds of moments from a statistical point of view concerning the signal and present relevant information as to different cardiovascular health parameters. The one employed in the suggested research study falls under the decision-making statistical feature extraction analysis that stands in the signal processing domain and is used to summarize the properties of various time-series data including the PCGs.

Table 3.1: Explanation of Statistical Features

Feature	Description
Mean	The average value of the signal
Standard Deviation (Std)	Shows how much the values vary from the mean
Minimum (Min)	The smallest value in the signal
Maximum (Max)	The highest value in the signal
Median	The middle value when the signal values are ordered
25th Percentile (Q25)	The value below which 25% of the data points fall
75th Percentile (Q75)	The value below which 75% of the data points fall
Range	The difference between the maximum and minimum values
Skewness	Measures if the values are more spread out on one side of the mean than the other
Kurtosis	Measures if the values are more peaked or flat compared to a normal distribution
Centroid	The center of the signals frequency components.
Root Mean Square (RMS)	The effective value of the signal, showing its power.
Chroma STFT	Analyzes the pitch classes in the signal, useful for identifying rhythms and patterns

3.5 Model Development

This step focuses on the creation and training of the ML classifiers relying on the PhysioNet 2016 data set. Important statistical characteristics from the dataset for training the several types of classifiers, namely, RF, SVM, and KNN. The proposed research supports the interaction of applying rigid feature extraction algorithms to maximize the dataset's signal aspects. This approach empowers the classifiers to detect and interpret patterns that signify diverse health conditions effectively. Figure 3.3 outlines the algorithm underlying the model development process, whereas Figure 3.4 provides a flowchart representation of the overall workflow.

ML-Based Model Development and Classification Algorithm

Step I: Initialization

Loaded the PhysioNet 2016 dataset for CVD classification.
Defined ML classifiers (Random Forest, SVM, KNN).
Set extracted feature list: mean, standard deviation, minimum, maximum, median, quartiles, range, skewness, kurtosis, centroid, RMS, Chroma STFT.

Step II: Feature Extraction

For each data sample:
Calculated statistical features to enhance dataset representation.
Assigned extracted features to classifiers for further processing.

Step III: Model Training

For each classifier (RF, SVM, KNN):
Trained classifiers with extracted features.
Optimized hyperparameters (e.g., kernel for SVM, decision trees for RF).
Stored trained model parameters for evaluation.

Case 1: Model Performance Evaluation

Step IV: Classification

For each trained model:
Classified PCG signals based on extracted features.
Applied RF, SVM, KNN

Step V: Model Selection and Evaluation

Evaluated each classifier using metrics:
Accuracy, precision, recall, specificity, F1 score.
Selected the best-performing model based on evaluation scores.

Case 2: Interpretability Analysis

Step VI: Model Interpretability Analysis

For the selected RF model:
Applied SHAP values and LIME to explain feature impact and measure feature's influence in the decision-making process.

Figure 3.3: Proposed ML-Based Model Development and Classification Algorithm

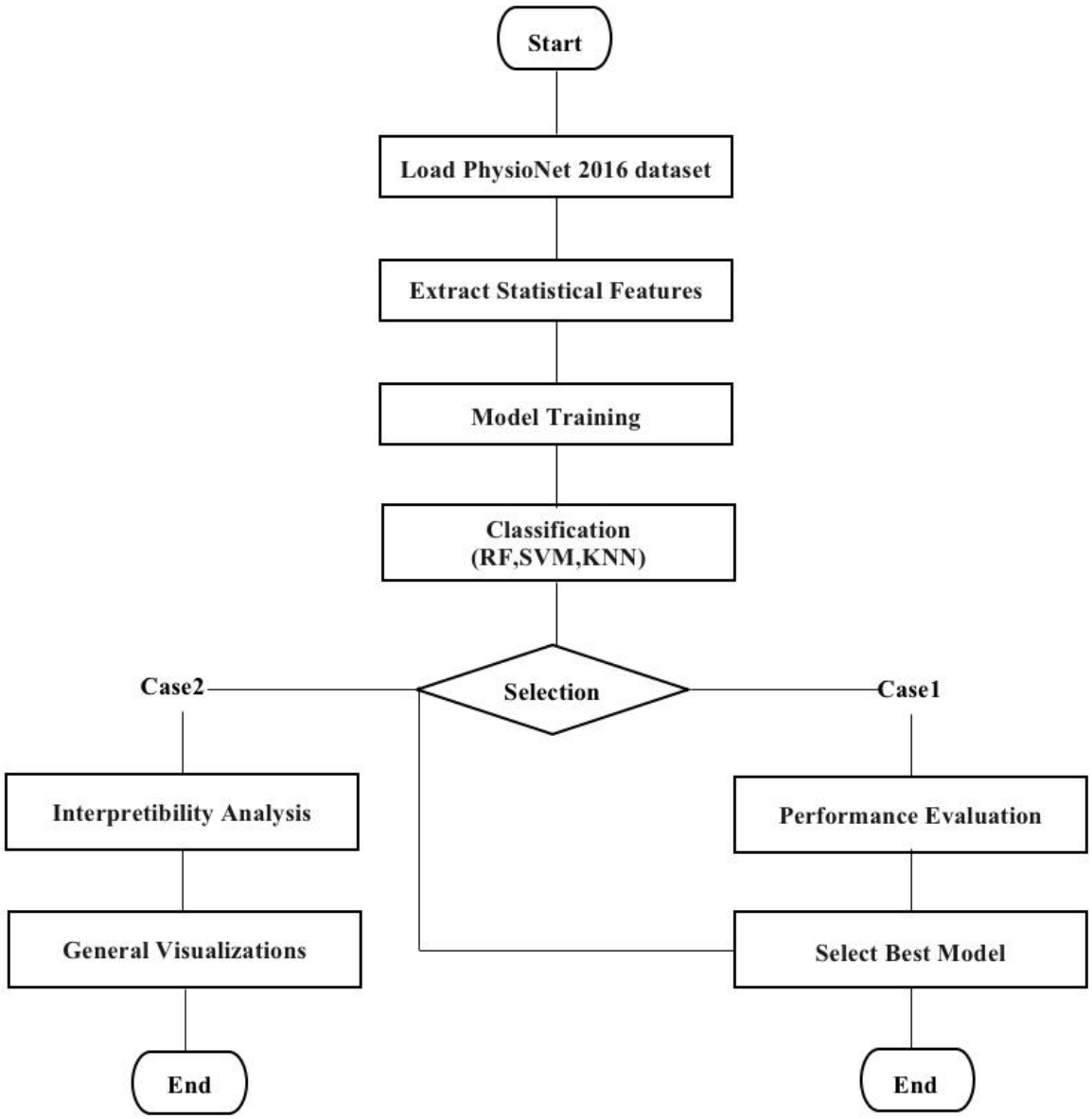

Figure 3.4: Workflow of the Model Development and Evaluation Process

3.6 Classification

In this step, the proposed research feeds the identified ML classifiers with the PCG signals to sort the signals based on the type of CVD. These classifiers are prepared by feeding the extracted features comprising of mean, standard deviation, minimum, maximum, median,

quartiles, range, skewness, kurtosis, centroid, RMS, and Chroma STFT to produce accurate and reliable results.

Finally, various XAI techniques are added to explain the outcomes that are obtained through various ML classifiers. The last step is to apply SHAP and LIME to measure the influence of features concerning the model's decision-making process. Besides providing details on the omission and selection of features to classify, SHAP and LIME enable the visualization of how the particular features impacted the entire model classification decision. This makes the model's decision to be more understandable, and the predictions more reliable and helpful to the clinicians because of compliance with clinical knowledge.

3.6.1 Model Selection

It is an important part of this research and aims at the elaborate comparison of several ML algorithms to identify the better-performing model for the classification of CVDs. The algorithms that were regarded are RF, SVM, and KNN classifiers.

Due to the classifier's power to correctly categorize data by finding the hyperplane that best differentiates the various classes, the SVM algorithm is selected. This process also requires changing some parameters like the kernel function and the regularization to boost the classification and minimize overfitting. Due to the random training samples and features, and the ability to acquire higher accuracy and low variance with multiple decision trees, RF is adopted for its ensemble learning. Tuning of the parameters is performed to enhance the predictor's accuracy and minimize the issue of overfitting. K-Nearest Neighbors (KNN) is employed because of its decision-making and easy explanation process. It focuses on the improvement of neighbors as well as distance measures for the successful classification of its result.

3.6.2 Model Interpretability

Throughout the given work, model interpretability is one of the main objectives and is attained through the use of explainability methods including SHAP (Shapley Additive explanations) and LIME (Local Interpretable Model-agnostic explanations). These techniques are also central in enlightening one on the preconditions for classification choices. Thus, by providing more definite estimates, clinicians and researchers can get a better understanding and trust the predictive power of the ML models used. It is an important phase in the proposed work, where only ML techniques are used for the proper classification of CVDs.

In the proposed research, it will be therefore important to adopt techniques from the explainability of machine learning to improve understanding of the results by the models and increase confidence. In particular, the SHAP technique is used to explain the Random Forest classifier's decision further to well-interpret the results. The SHAP values point out the contribution each feature has to the model predictions, something that is very useful to clinicians and researchers. This approach enhances interpretability and can be utilized to explain an ML model's outputs concerning CVD classification.

3.7 Model Evaluation

In this step, the effectiveness and reliability of the developed models for classifying CVDs using XAI based on PCGs are rigorously assessed. Various metrics and analyses are employed to ensure the models meet the desired standards of different parameters against potential biases or errors.

3.7.1 Performance Evaluation Metrics

They are widely used forms of measures that can aid in quantitatively evaluating the efficiency of the generated models in CVD classification. Various measurements like accuracy, precision, recall, specificity, and F1 Score were calculated to measure the model's predictive efficiency and its uttermost performance. The aforementioned metrics give information regarding the possibility of the distinct classification of different cardiovascular conditions with the help of PCG signals and, therefore, aid in decision-making regarding the further improvement and enhancement of the models.

3.7.2 Interpretability Analysis

Regarding the choice of interpretability analysis, it is worth emphasizing that Interpretability analysis is to assess how well the applied XAI techniques help in understanding the models decision-making process. Concerning model interpretation, this analysis uses methods like SHAP and LIME. This section focuses on providing an improved understanding of model interpretability and offering information about the AI-based diagnostic instruments for CVDs that would satisfy medical practitioners and meet all the regulatory necessities.

CHAPTER 4

SIMULATION RESULTS AND DISCUSSION

4.1 Overview

As for the results of the simulation based on the use of ML classifiers on the PhysioNet 2016 dataset, this part provides a detailed discussion of the findings. Strictly discuss, the performance comparison of the used classifiers, RF, KNN, and SVM, is also a part of it. In addition, the ways to improve the interpretability of the model are described through the use of the SHAP (Shapley Additive explanations) and LIME (Local Interpretable Model-agnostic explanations). LIME and SHAP were chosen because they effectively explain machine learning models, especially in healthcare. LIME gives clear, detailed insights into individual predictions, helping clinicians understand specific cases. SHAP provides a consistent view of how each feature impacts predictions based on solid theory. Both tools are well-documented and widely used, making them reliable choices for enhancing trust and transparency in model interpretations.

4.2 Experimental Setup

The proposed study employed the PhysioNet 2016 dataset, which consists of both healthy and unhealthy phonocardiogram (PCG) signals. The dataset originally contained 659

instances labeled as patients and 3,290 instances labeled as healthy, highlighting a significant class imbalance. To mitigate this issue, we implemented the Synthetic Minority Over-sampling Technique (SMOTE), which balanced the classes by creating synthetic samples for the minority class, resulting in an equal number of instances for both categories after preprocessing. The preprocessing steps included signal denoising using a fourth-order Butterworth bandpass filter, ensuring the retention of critical frequency components between 20 Hz and 400 Hz.

For model development, we utilized various machine learning classifiers, specifically Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF). The classifiers were trained using the extracted statistical features from the PCG signals, which included metrics such as mean, standard deviation, and skewness, among others. The model performance was evaluated using accuracy, precision, recall, specificity, and F1 score. All simulations were conducted using Python with libraries such as scikit-learn and imbalanced-learn on the Apple MacBook Air. Table 4.1 summarizes the key simulation parameters used in this study.

Table 4.1: Simulation Parameters

Parameter	Value
Dataset	Physionet 2016 Dataset
Original Number of Samples	3949 (659 patients, 3290 healthy)
Post-SMOTE Samples	3290 (balanced for both classes)
Feature Extraction Method	Statistical Feature Extraction
Signal Denoising Method	Butterworth bandpass filter (20-400 HZ)
Train-Test Split	85% Training, 15% Testing
SVM Kernel	Polynomial
KNN Neighbors	5
Random Forest Estimators	100
Standardization Method	StandardScaler
Evaluation Metrics	Accuracy, Precision, Recall, Specificity, F1 Score
Software Used	Python, Scikit-learn, imbalanced-learn
Hardware	Apple MacBook Air

4.3 Classification Results

The major Machine Learning classifiers which are prevalently used are assessed based on their efficiency.

4.3.1 Random Forest (RF)

The RF classifier achieved an accuracy of 93.82%, precision of 92.01%, recall of 95.33%, specificity of 92.44%, and an F1 score of 93.64%. These results indicate that the RF model performs exceptionally well in classifying heart sounds, balancing both precision and recall effectively.

4.3.1.1 Confusion Matrix Analysis of RF Classifier

It is possible to assess the performance of the RF classifier using a confusion matrix in detail. The confusion matrix offers an understanding of the efficiency of the classifier as it gives the exact number of correct and wrong predictions regarding classes. As for the confusion matrix of the RF classifier of this study, it is presented in Figure 4.1.

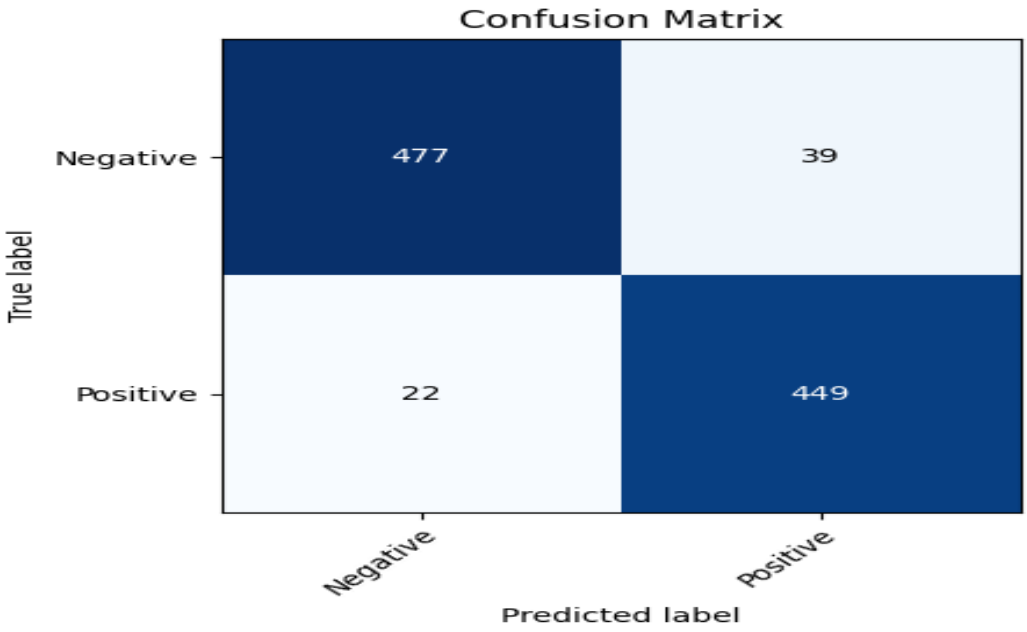

Figure 4.1: CM of RF Classifier

The metrics representing the confusion matrix to the RF classifier concerning the classification of CVDs are shown in Table 4.2. It gives the number of TP, TN, FP, and FN that would enable the RF model to show how nicely it can differentiate between positive and negative CVD patients. The percentage of each component is calculated by using Equation 4.2 to Equation 4.5.

$$\text{Total} = TP + TN + FP + FN \quad (4.1)$$

$$\text{Total} = 449 + 477 + 39 + 22$$

$$\text{Total} = 987$$

$$\text{TP\%} = \frac{TP}{Total} \times 100 \quad (4.2)$$

$$\text{TP\%} = \frac{449}{987} \times 100 = 45.49\ \%$$

$$\text{TN\%} = \frac{TN}{Total} \times 100 \quad (4.3)$$

$$\text{TN\%} = \frac{447}{987} \times 100 = 48.33\ \%$$

$$FP\% = \frac{FP}{Total} \times 100 \quad (4.4)$$

$$FP\% = \frac{39}{987} \times 100 = 3.95\ \%$$

$$FN\% = \frac{FN}{Total} \times 100 \quad (4.5)$$

$$FP\% = \frac{22}{987} \times 100 = 2.23\ \%$$

Table 4.2: Breakdown of the (RF) Confusion Matrix

Category	Description	Percentage (%)
TP	Actual cases positive (CVDs), correctly predicted positive by the model	45.49
TN	Actual cases negative (no CVDs), correctly predicted negative by the model	48.33
FP	Actual cases negative, mistakenly predicted positive by the model	3.95
FN	Actual cases positive, mistakenly predicted negative by the model	2.23

Metrics Derived from the Confusion Matrix (RF)

From the confusion matrix, several important performance metrics such as accuracy, precision, recall, F1 score, and specificity are derived for the RF classifier as shown in Equation 4.6 to Equation 4.9 [41] and Equation 4.10 [10].

Accuracy: The overall correctness of the model

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.6)$$

$$= \frac{449+477}{449+477+39+22} = 0.9382$$

Precision: The amount of correct positive predictions

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4.7)$$

$$= \frac{449}{449+39} = 0.9201$$

Recall: The number of actual positives that are accurately identified

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4.8)$$

$$= \frac{449}{449+22} = 0.9533$$

F1 Score: The harmonic mean of precision and recall

$$\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4.9)$$

$$= 2 \times \frac{0.9201 \times 0.9533}{0.9201 + 0.9533} = 0.9364$$

Specificity: The amount of actual negatives that are accurately identified

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (4.10)$$

$$= \frac{477}{477+39} = 0.9244$$

These metrics collectively indicate that the RF classifier performs exceptionally well in distinguishing between the presence and absence of CVDs using PCGs. The high precision, recall, and F1 score signify the model's robustness, formulating it a trustworthy tool for clinical diagnostics.

4.3.2 K-Nearest Neighbours (KNN)

The KNN classifier yielded an accuracy of 86.52%, precision of 82.50%, recall of 91.08%, specificity of 82.36%, and an F1 score of 86.58%. While slightly lower than RF, KNN still demonstrates a strong performance, particularly in the recall, which indicates its effectiveness in identifying positive cases.

4.3.2.1 Confusion Matrix Analysis of KNN Classifier

The performance of the KNN classifier can be evaluated using the confusion matrix as shown in Figure 4.2.

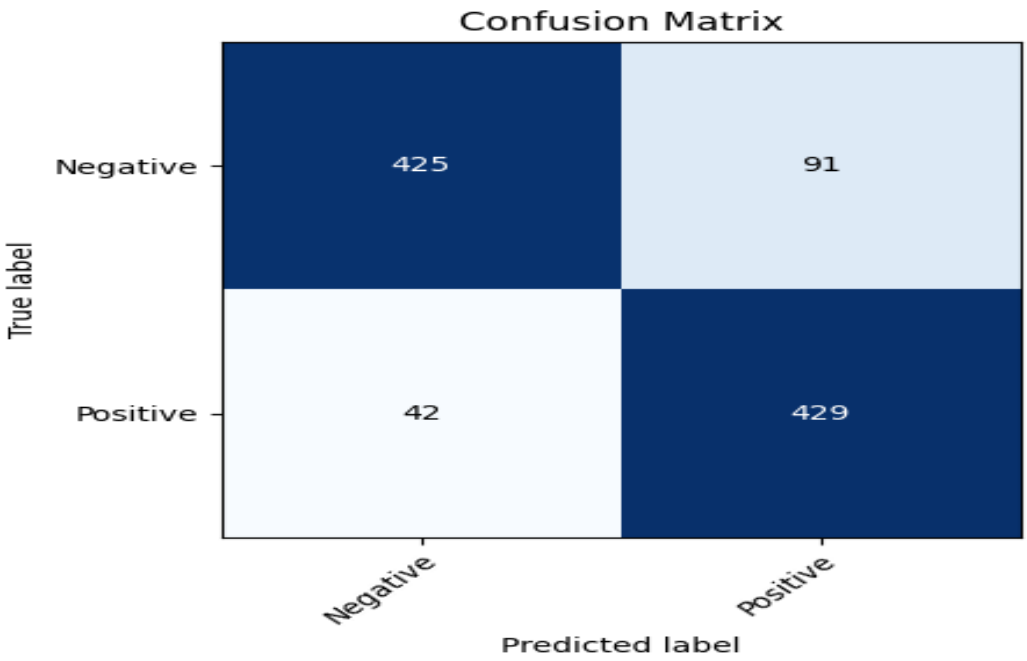

Figure 4.2: CM of KNN

Table 4.3 presents the metrics of the KNN classifier for the classification of CVDs. It details the counts of TP, TN, FP, and FN, illustrating how effectively the KNN model distinguishes between positive and negative CVD cases. The percentage of individually component is regarded by applying Equation 4.2 to Equation 4.5.

Total = $429 + 425 + 91 + 42 = 987$

TP% = $\frac{429}{987} \times 100 = 43.45\ \%$

TN% = $\frac{425}{987} \times 100 = 43.07\ \%$

FP% = $\frac{91}{987} \times 100 = 9.22\ \%$

FP% = $\frac{42}{987} \times 100 = 4.25\ \%$

Table 4.3: Breakdown of the (KNN) Confusion Matrix

Category	Percentage (%)
TP	43.45
TN	43.07
FP	9.22
FN	4.25

Metrics Derived from the Confusion Matrix (KNN)

From the confusion matrix in Figure 4.2, we can derive several important performance metrics for the KNN classifier by using Equation 4.6 to Equation 4.10:

Accuracy = $\frac{429+425}{429+425+91+42} = 0.8652$

Precision = $\frac{429}{429+91} = 0.8250$

Recall = $\frac{429}{429+42} = 0.9108$

F1 Score = $2 \times \frac{0.8250 \times 0.9108}{0.8250+ 0.9108} = 0.8658$

Specificity = $\frac{425}{425+91} = 0.8236$

These metrics provide an inclusive evaluation of the KNN classifier performance in classifying CVDs using PCGs. While the model demonstrates good accuracy and recall, indicating its capability to identify both positive and negative cases effectively, there is room for improvement in reducing false positives and false negatives to enhance diagnostic accuracy.

4.3.3 Support Vector Machine (SVM)

This classifier resulted in an accuracy of 77.81%, precision of 75.40%, recall of 79.41%, specificity of 76.36%, and an F1 score of 77.35%. Although SVM showed the lowest performance among the three classifiers, it provides a baseline for comparison and highlights areas for potential improvement.

4.3.3.1 Confusion Matrix Analysis of SVM Classifier

The performance of the SVM classifier can be evaluated using the confusion matrix as shown in Figure 4.3.

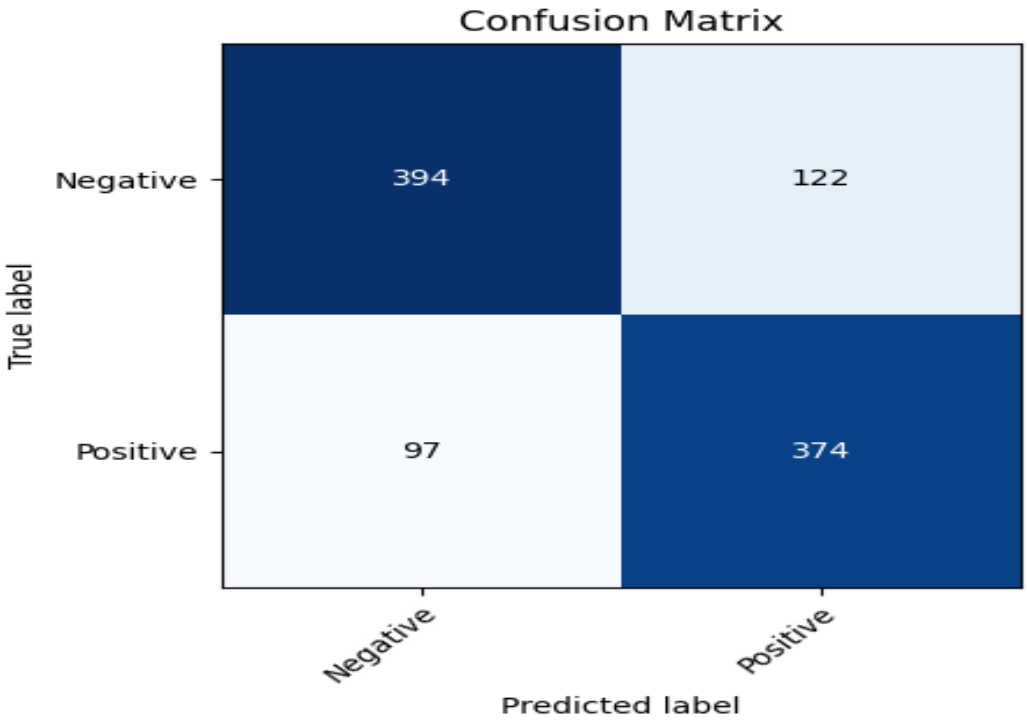

Figure 4.3: CM (SVM)

Table 4.4 presents the metrics of the SVM classifier for the classification of CVDs. It details the counts of TP, TN, FP, and FN, illustrating how effectively the SVM model distinguishes between positive and negative CVD cases. It is considered by using a percentage of individually component by employing Equation 4.2 to Equation 4.5.

$$\text{Total} = 374 + 394 + 122 + 97 = 987$$

$$\text{TP\%} = \frac{374}{987} \times 100 = 37.89\ \%$$

$$\text{TN\%} = \frac{394}{987} \times 100 = 39.93\ \%$$

$$\text{FP\%} = \frac{122}{987} \times 100 = 12.36\ \%$$

$$\text{FP\%} = \frac{97}{987} \times 100 = 9.83\ \%$$

Table 4.4: Breakdown of the (SVM) Confusion Matrix

Category	Percentage (%)
TP	37.89
TN	39.93
FP	12.36
FN	9.83

Metrics Derived from the Confusion Matrix (SVM)

From the confusion matrix in Figure 4.3, we can derive several important performance metrics for the SVM classifier by using Equation 4.6 to Equation 4.10:

$$\text{Accuracy} = \frac{374+394}{374+394+122+97} = 0.7781$$

$$\text{Precision} = \frac{374}{374+122} = 0.7540$$

$$\text{Recall} = \frac{374}{374+97} = 0.7941$$

$$\text{F1 Score} = 2 \times \frac{0.7540 \times 0.7941}{0.7540 + 0.7941} = 0.7735$$

$$\text{Specificity} = \frac{394}{394+97} = 0.7636$$

These metrics provide a broad evaluation of the SVM classifier performance in classifying CVDs using PCGs. The model demonstrates moderate accuracy, precision, recall, specificity, and F1 score, indicating its potential utility in clinical applications but also highlighting areas for improvement in diagnostic accuracy, especially in reducing false positives and false negatives.

4.4 Comparison of Classifier Performance

The performance metrics of the three classifiers are summarized in Table 4.5. RF outperformed both KNN and SVM in all metrics, making it the most robust classifier for this dataset. KNN showed commendable performance, particularly in the recall, whereas SVM lagged in most metrics but still provided valuable insights.

Table 4.5: Performance Comparison of RF, KNN, SVM

Classifier	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 Score (%)
RF	93.82	92.01	95.33	92.44	93.64
KNN	86.52	82.50	91.08	82.36	86.58
SVM	77.81	75.40	79.41	76.36	77.35

4.5 Model Interpretability

Model interpretability is crucial for accepting the decision-making process of machine learning models, especially in clinical settings. SHAP and LIME values were utilized to interpret the results of the best-performing model, RF, by determining the impact of each feature on the model's predictions. A dependency graph is a visual representation that illustrates the relationships between different variables or features within a dataset. Each node in the graph represents a variable, while directed edges indicate dependencies or influences between them. In the context of machine learning, particularly in interpretability methods like SHAP and LIME, dependency graphs help us understand how specific features interact with each other and contribute to model predictions. By analyzing these dependencies, researchers can gain insights into the underlying mechanisms driving the model's decisions, which is particularly important in high-stakes fields such as healthcare.

4.5.1 SHAP Analysis

It provides a unified measure of feature importance, enabling detailed insights into how individual features influence the model's predictions. Figures 4.4 through figure 4.20 display the summary plot, dependence plot, Bar plot, Waterfall plot, and force plot, respectively.

4.5.1.1 SHAP Summary Plot

In this plot, features are ranked by their average SHAP values showing the most important features at the top and the least important ones at the bottom. This helps to understand the impact of each feature on the model's predictions. [42]

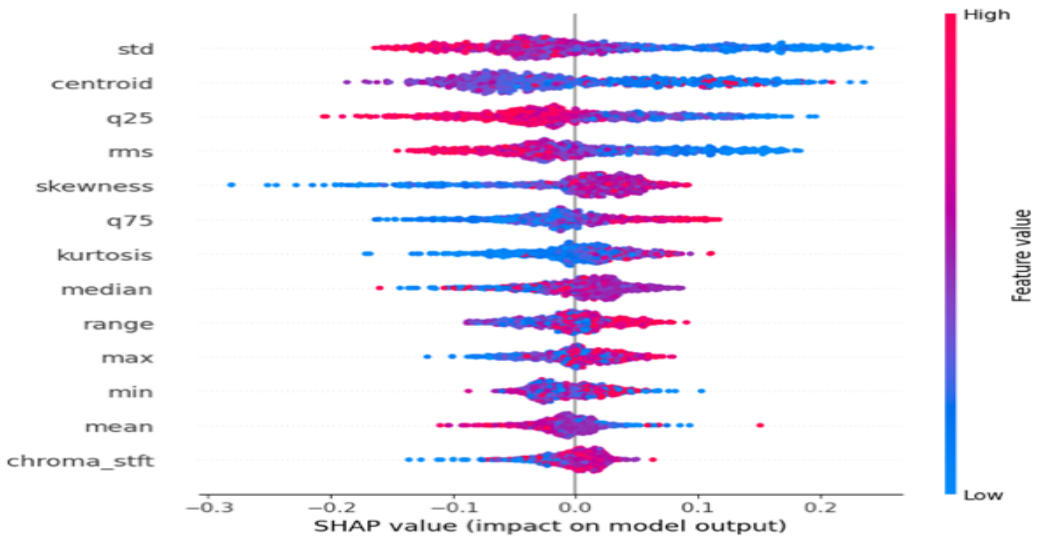

Figure 4.4: Summary Plot

Figure 4.4 demonstrates the impact of various features on the model's prediction, visualizing both the magnitude and direction of their effects. The x-axis represents the SHAP

values, which quantify the influence of each feature on the model output, ranging approximately from -0.3 to 0.2. The y-axis lists the features under consideration, such as rms, centroid, std, and q25. The spread of the dots along the x-axis indicates the variance in the impact of each feature. The most significant feature lies at the top of all features in this study ‘std’ is the most significant feature. The feature ‘rms’ exhibits a wide spread of SHAP values, suggesting it has also a significant and varied influence on the model’s predictions. The color coding of the dots indicates low or high feature values. Blue dots indicate low feature values and red dots indicate high feature values. It also reveals the relationship between the feature values and their SHAP values. High rms values or red dots are generally associated with positive SHAP values, indicating that higher rms values increase the model’s prediction.

Similarly, the feature std shows that low values (blue dots) tend to have negative SHAP values, thus reducing the prediction, whereas high values (red dots) increase the prediction. Features such as q25 and q75 also demonstrate this trend, where higher values push the prediction higher. In contrast, the feature min has SHAP values close to zero, representing a minimal impact on the model’s predictions compared to other features. Furthermore, features like kurtosis and mean display a balanced spread of SHAP values, with high values significantly contributing to positive predictions. The chroma_stft feature similarly shows that high values increase the model output, while low values decrease it. Features such as rms, std, q25, and q75 are shown to have substantial impacts, as indicated by the broad spread and color gradation of their SHAP values. This analysis is crucial for understanding the underlying mechanisms of the model, ensuring transparency, and validating that the model's behavior aligns with domain knowledge.

4.5.1.2 SHAP Dependence Plot

Unlike summary plots, dependence plots show the relationship between a specific feature and the predicted outcome for each instance within the data. This analysis is performed for multiple reasons and is not limited to gaining more granular information and validating the importance of the feature being analyzed by confirming or challenging the findings from the SHAP summary plots or other global feature importance measures. [42]

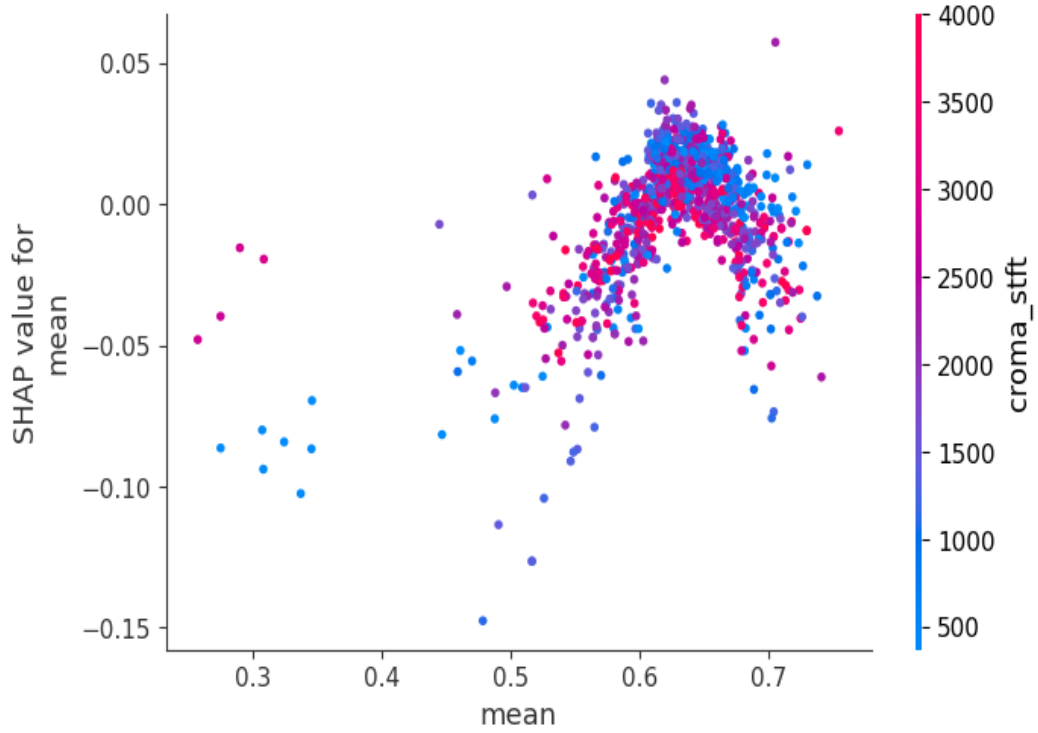

Figure 4.5: Dependence Plot (Mean)

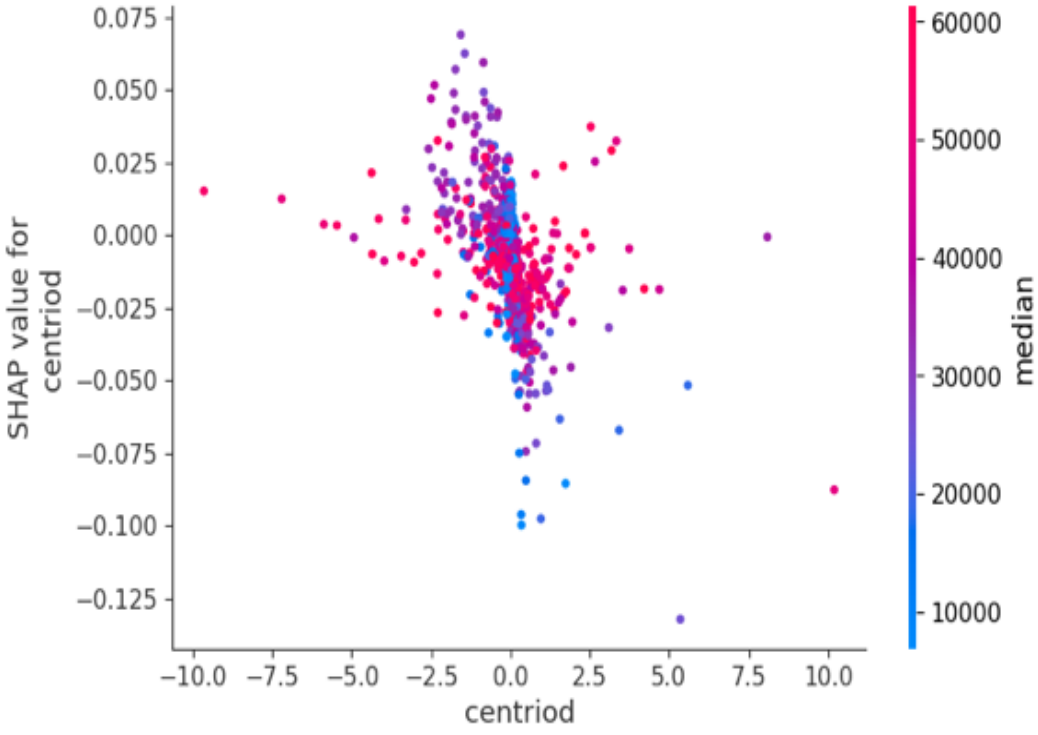

Figure 4.6: Dependence Plot (Centriod)

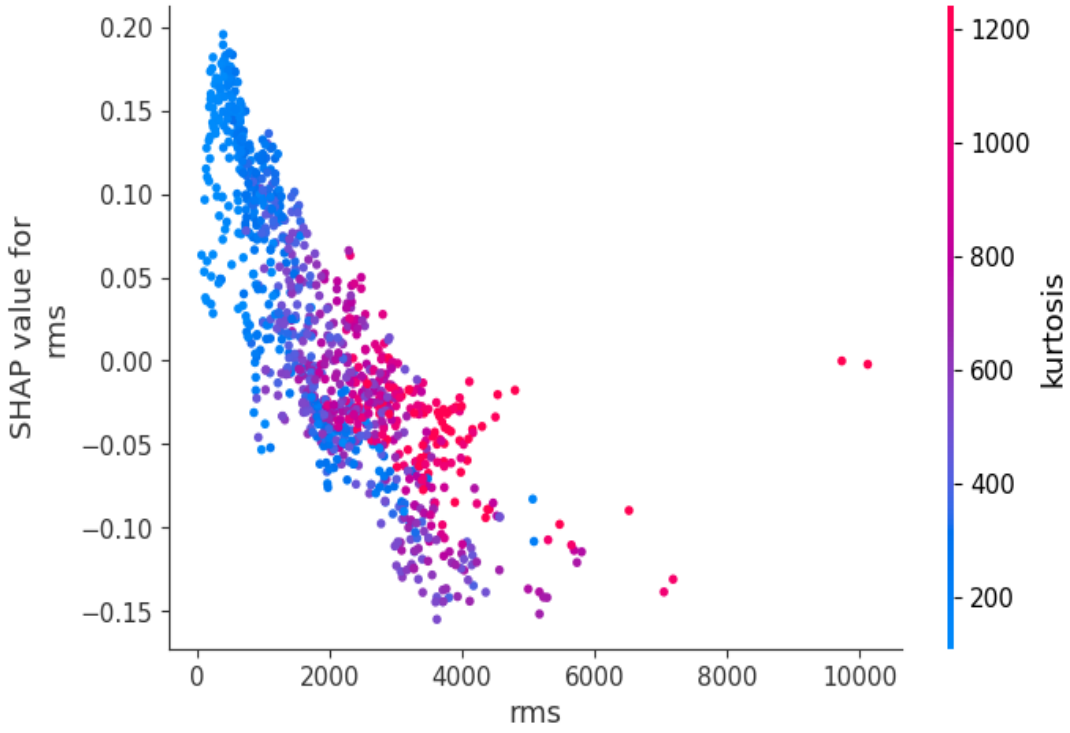

Figure 4.7: Dependence Plot (rms)

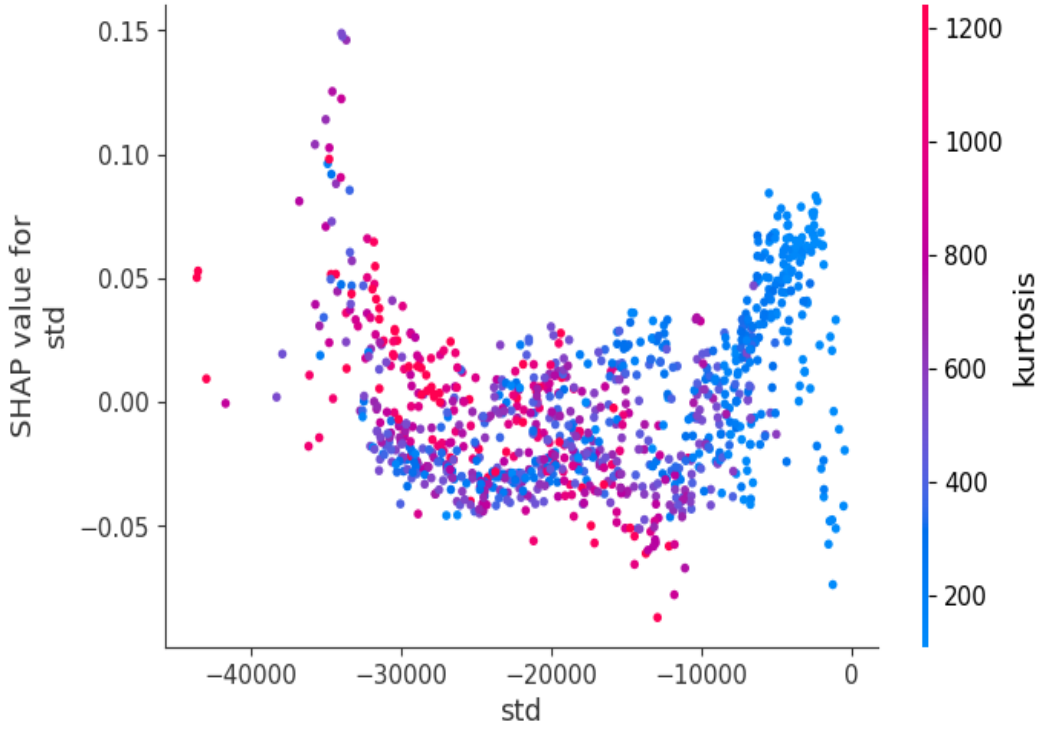

Figure 4.8: Dependence Plot (std)

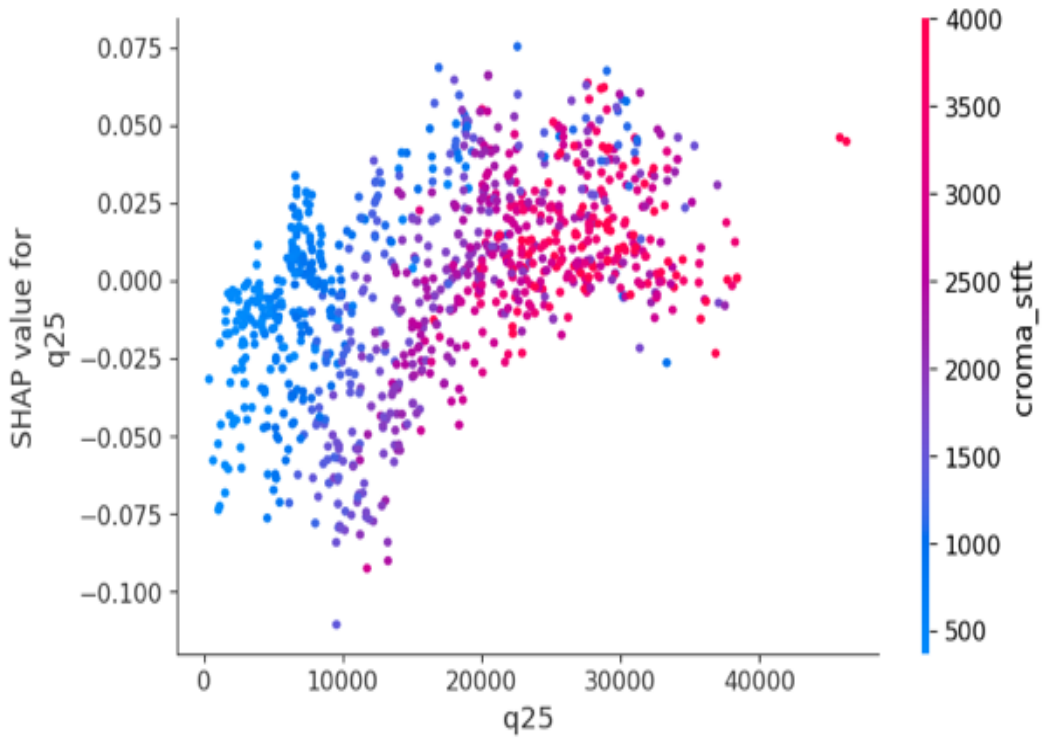

Figure 4.9: Dependence Plot (q25)

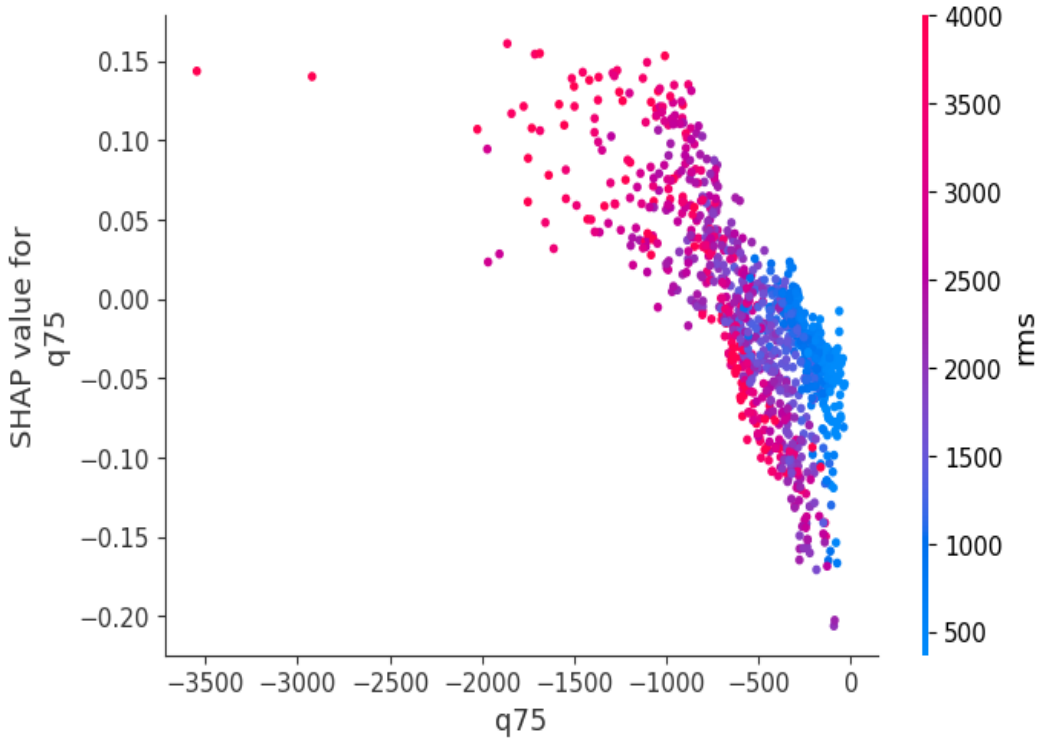

Figure 4.10: Dependence Plot (q75)

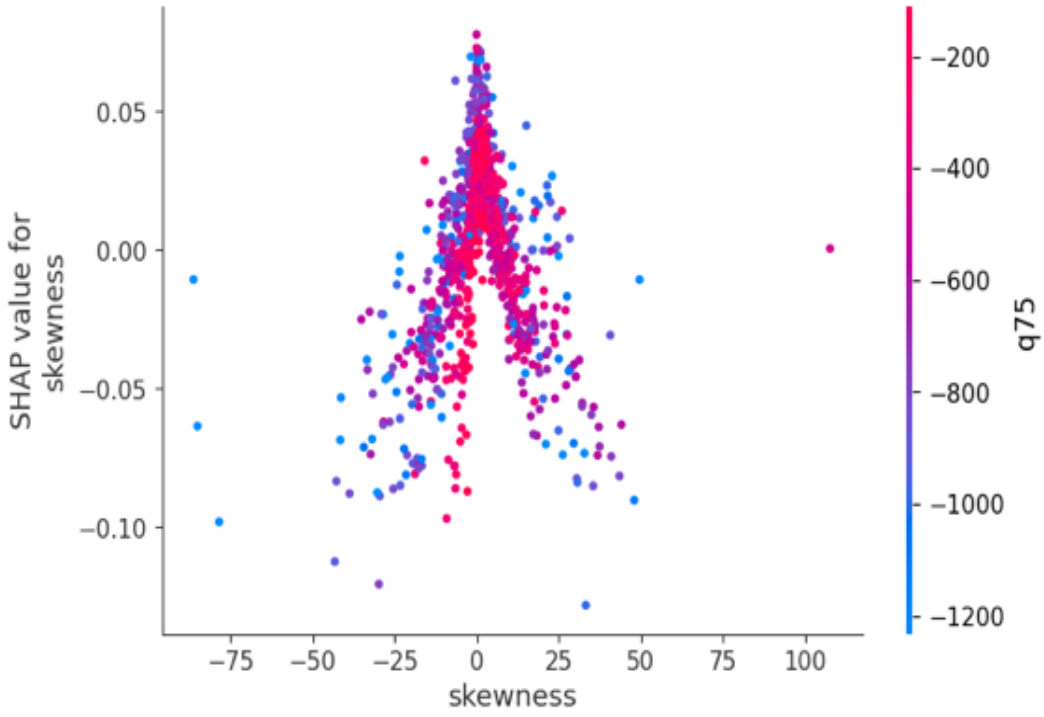

Figure 4.11: Dependence Plot (Skewness)

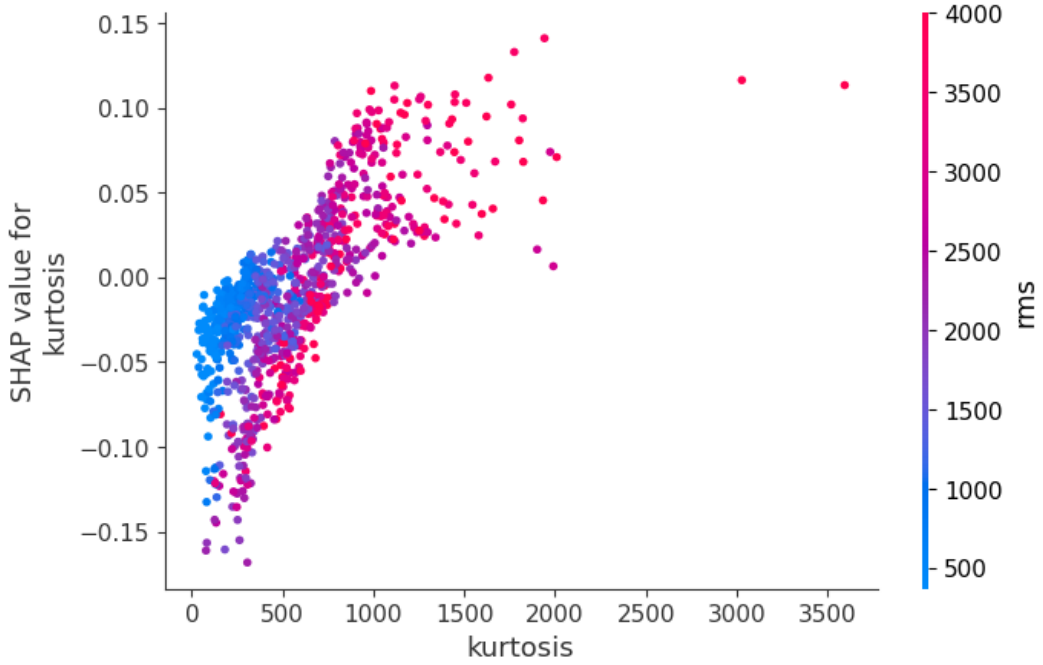

Figure 4.12: Dependence Plot (Kurtosis)

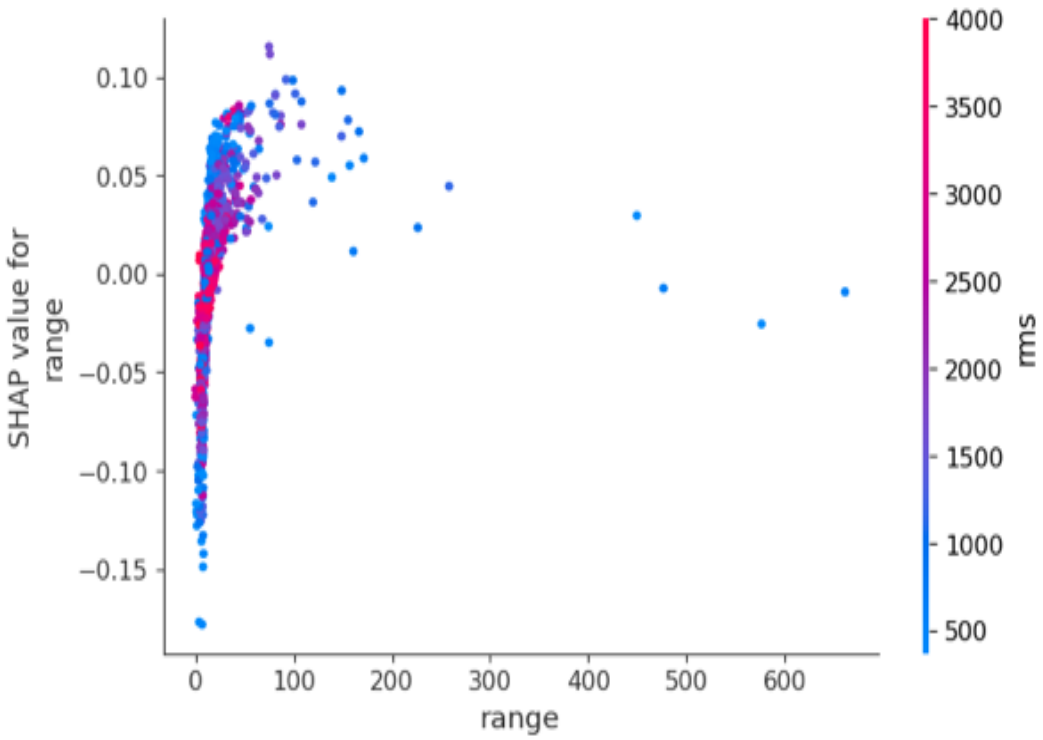

Figure 4.13: Dependence Plot (Range)

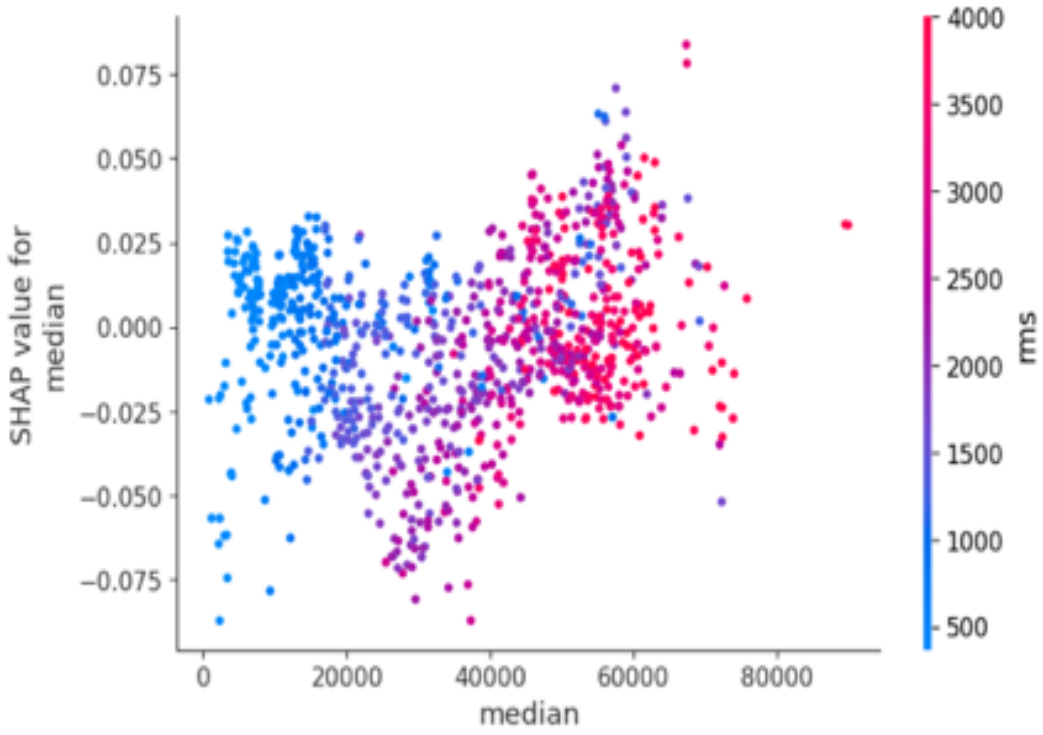

Figure 4.14: Dependence Plot (Median)

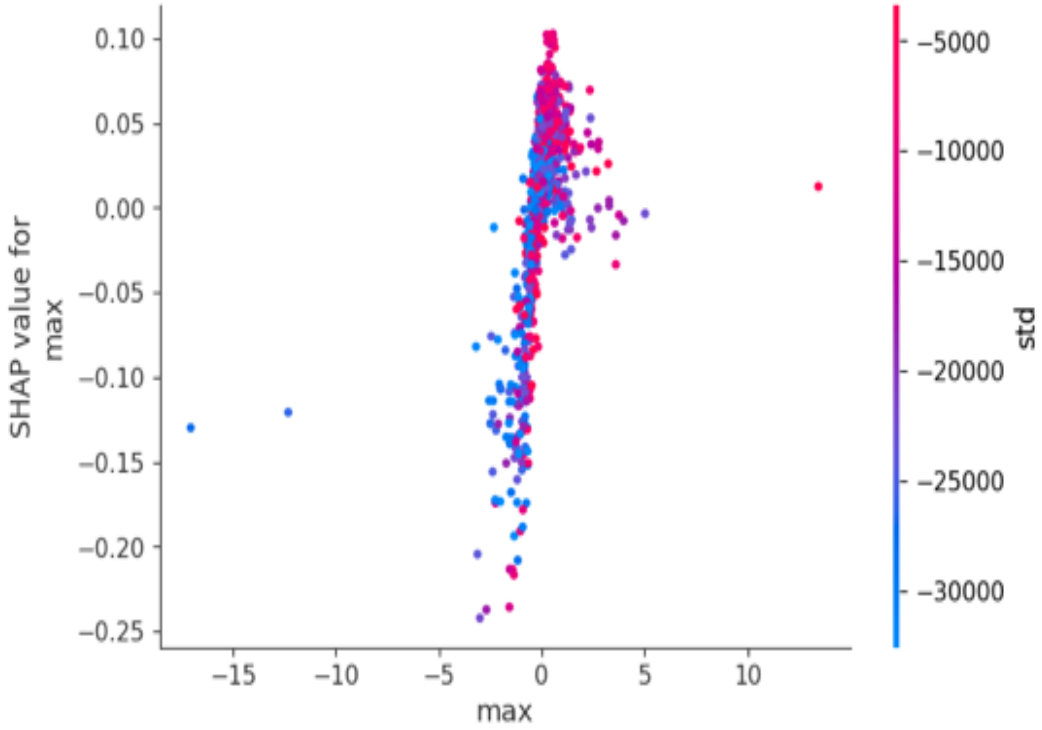

Figure 4.15: Dependence Plot (Max)

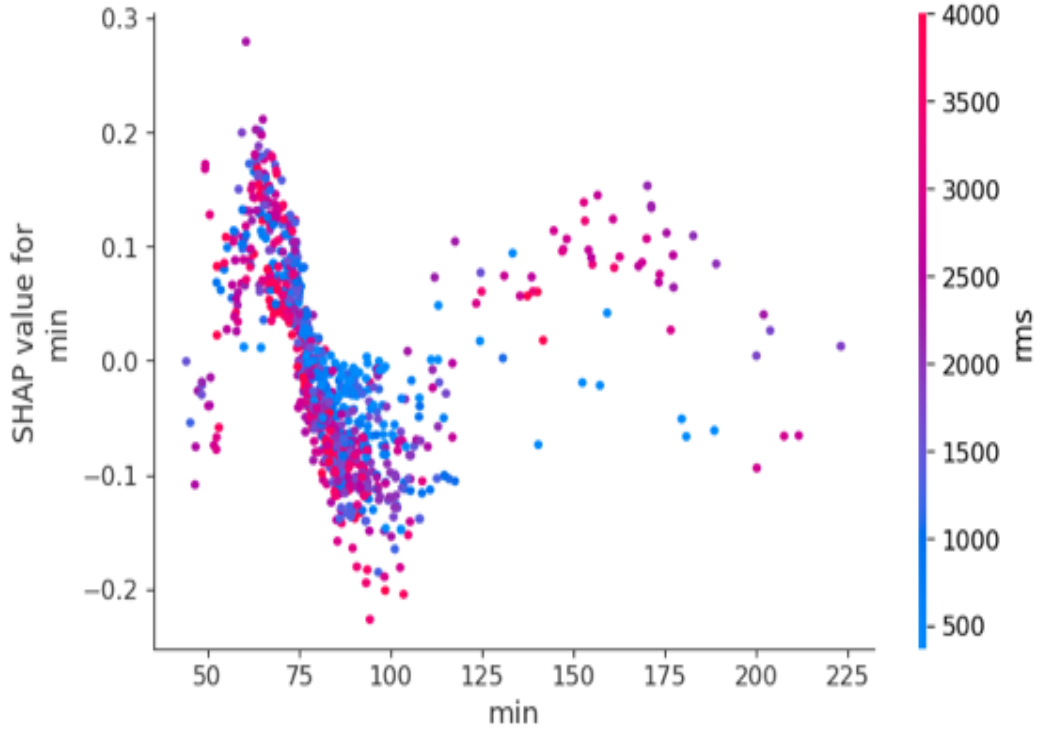

Figure 4.16: Dependence Plot (Min)

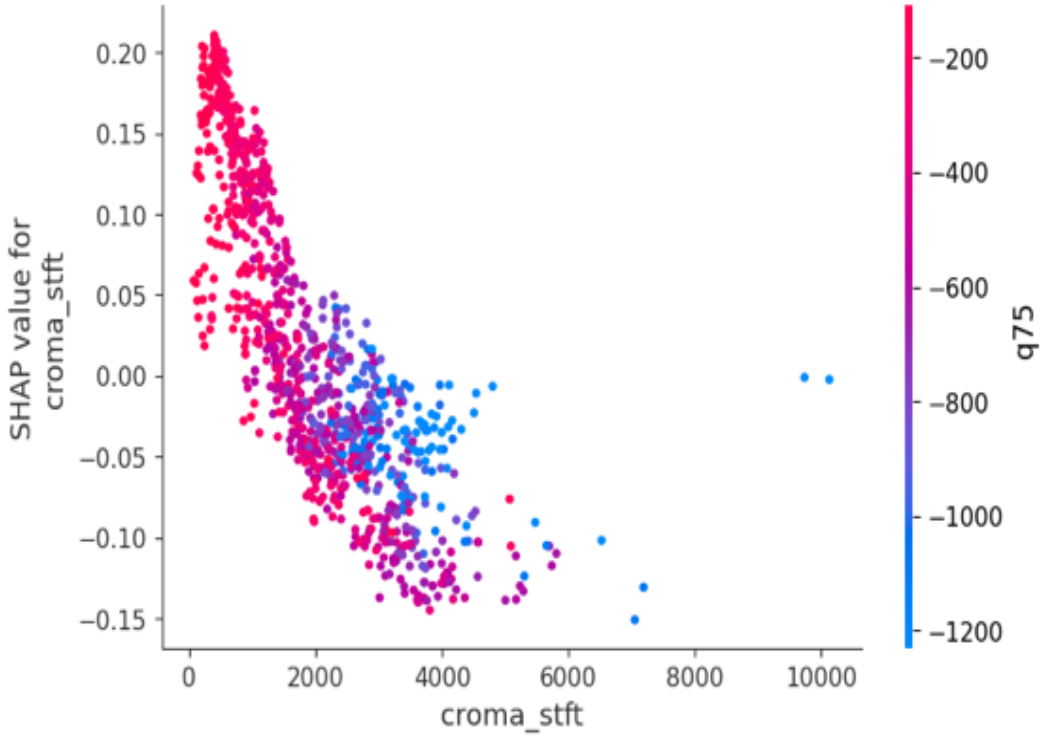

Figure 4.17: Dependence Plot (Croma_stft)

The analysis of the SHAP dependence plot as shown in Figure 4.5 to Figure 4.17 reveals several key insights into the relationship between the features 'mean' and 'chroma_stft' and their impact on the model predictions. From Figure 4.5 to Figure 4.17, the x-axis represents the mean feature values, while the y-axis shows the SHAP values for the mean feature. The color gradient indicates the values of features, with blue representing lower values and red representing higher values. There is a noticeable trend where lower mean values tend to have negative SHAP values, implying a decrease in the model's prediction. As the mean values increase, the SHAP values generally move towards zero or slightly positive, indicating a neutral or mildly positive impact on the prediction. This suggests that higher mean values are associated with either no change or a slight increase in the likelihood of the predicted outcome.

From Figure 4.5 to Figure 4.17, the opaqueness of dots also underlines the interdependence of characteristics, for example in the case of figure (a) 'mean' and 'chroma_stft'. Representing the higher feature values there are red dots which are located at

the higher means, whereas, the lower feature values are depicted by the blue dots which are spread out and hence have lower means. This pattern suggests that higher values of 'chroma_stft' occur where the 'mean' feature is also higher and it can be thought that this helps to positively alter the model's prediction.

Looking at the location measures, the majority of data points are closely grouped around the mid-point or mean value of 0.55 to 0.65 where the SHAP values are near zero. Based on this clustering, we can infer that within the range of mean values, the feature contributes a minimal extent to causing modification in the model's prediction, and perhaps possesses a threshold value. This threshold effect could suggest that within this range the 'mean' feature is not effective in changing the output of model and thus should be explored in detail. In totality, these observations are rather beneficial when determining how the features interconnect or how a certain feature impacts the model's prediction about others.

4.5.1.3 SHAP BAR Plot

It is a type of visualization that reveals features contribution in the machine learning model. Here, each bar in the plot corresponds to a feature, and the length of each bar signifies how much of the model's decision was influenced by the respective feature. The below plot is useful in model interpretability, as it sorts features by the mean absolute SHAP value hence it is easy to understand the model decisions.

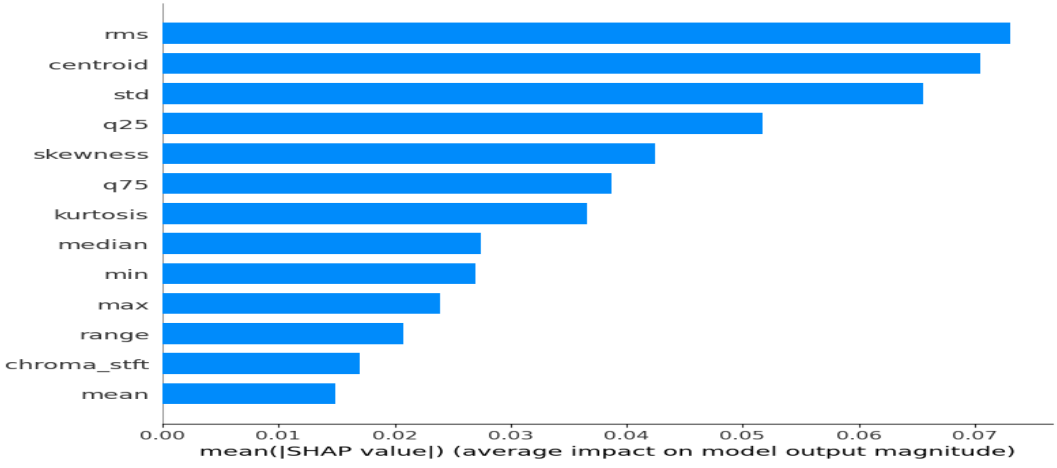

Figure 4.18: Bar Plot

The bar plot shown in Figure 4.18 illustrates the average absolute SHAP values for various features, thereby indicating their overall importance in the model's prediction. The x-axis represents the mean of the absolute SHAP values for each feature and the y-axis lists the features analyzed, including rms, centroid, std, and others. Key observations from this figure include the ranking of feature importance, with the rms feature having the highest mean absolute SHAP value, suggesting it has the most significant impact on the model predictions. The centroid and std features also show high mean absolute SHAP values, highlighting their importance in influencing the model's decisions. Features such as q25, skewness, and q75 follow, demonstrating moderate impacts on the model output. The magnitude of impact is particularly notable for rms, which has an average absolute SHAP value of approximately 0.07, signifying its strong influence on predicted outcomes.

The centroid feature is most impactful with an average absolute SHAP of the feature value around 0.06, which similarly indicates a very significant impact on prediction. The decreasing effect is observed for such features as std and q25, as mean absolute SHAP values are approximately 0.05 and 0.04, respectively. The mean absolute SHAP of range and stft features is comparatively minor, thus, implies range, stft, and mean are the features with the least impact on the model. For the selected features, their mean absolute SHAP values are approximately equal to 0.01 to 0.02, suggesting minimal impact. This bar plot can efficiently prioritize the features with the highest mean values and sort the features based on their contribution to the model's prediction since the higher the value, the more the feature affects its prediction This bar plot shows that rms, centroid, and std are the most influential features, whereas, mean, chroma_stft, and range are less important. This visualization is necessary for the analysis of which features are the most influential in the making of the final decision and for the exclusion of the situations that a data scientist may consider undesired from the model's decision process.

From the SHAP summary plot, it is noticed that the features with the largest decision importance are std, centroid, q25, and rms. This means these features play a significant role in the identified model when considering how and to what extent they impact the results. On the other hand, the SHAP bar plot orders the features from rms, to followed by centroid, std, and q25. This derives a conclusion that rms is the most influential feature based on its signification on the average influence of the model.

Others are the attributes that are identified in the model such as impact and variability, which relates to the contribution made by a feature to the model prediction, as indicated by SHAP values. The larger the SHAP value, the higher the contribution or impact on the prediction. In summary plots, those features with high influence are often listed first to indicate the importance of those features in the model's evaluation. Variability, on the other hand, can be understood as the variability of SHAP values for a certain feature within all the predictions. This implies that the feature has a large average causal effect on the predictions but with large variation in this effect implying that the variability or how this effect changes from instance to instance is important in prediction.

The distinctions between these plots result from the fact that they each denote a different variable. The summary plot also assesses the impact and variability and is wider in its perspective. The bar plot is based only on the impact concept which separates it into only the average influence of each feature and gives a basic ranking. Both plots are useful the summary plot provides details of the feature importance sequentially and thus provides more details, on the other hand, the bar plot straightforwardly depicts the ranking.

4.5.1.4 SHAP Waterfall plot

It is beneficial as it shows how each feature value affects the model's prediction in this instance, giving one good insight into how the model arrived at such a decision.

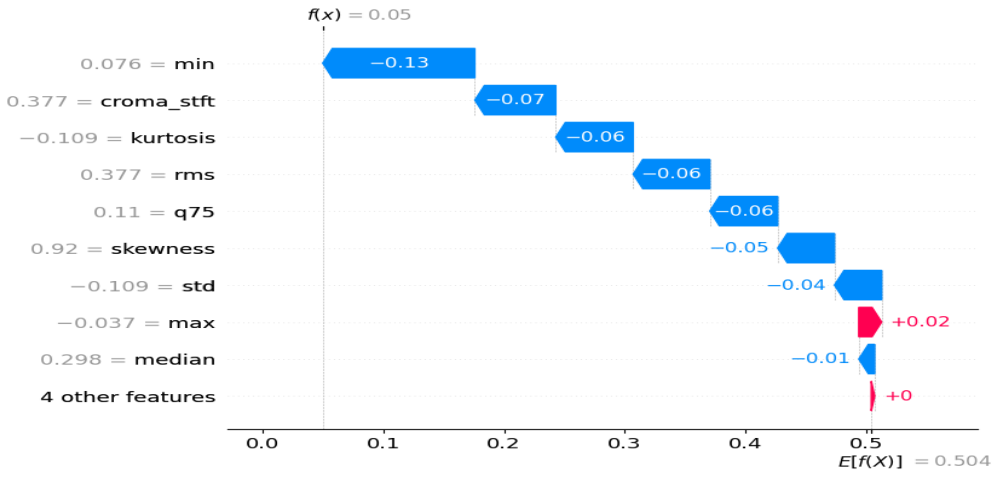

Figure 4.19: Waterfall Plot

Figure 4.19 shows the features involved in the breakdown of the particular instance prediction. The plot's starting point is the expected value (E[f(x)]) of 0. 504, which represents the average model output across all samples in the dataset. Every single point in the plot is associated with an effect of one of the features on the model's outcome. The features are labeled at the y-axis beginning from min, croma_stft, kurtosis, and others. The actual values of these features for the instance are indicated beside the feature names (e. g., 0. 076 for the min feature).

The horizontal bars in Figure 4.19 represent the contribution of each feature to the prediction, with the length and direction of the bars indicating the magnitude and direction of the contribution. Blue bars signify negative contributions, pushing the prediction lower, while red bars indicate positive contributions, pushing the prediction higher. For this instance, the feature 'min' with a value of 0.076 decreases the prediction by 0.13 units. Similarly, 'croma_stft' with a value of 0.377 decreases the prediction by 0.07 units, and 'kurtosis' with a value of -0.109 decreases it by 0.06 units. Other features such as 'rms', 'q75', 'skewness', 'std', and 'max' also decrease the prediction by varying amounts. The 'median' feature with a value of 0.298 slightly increases the prediction by 0.02 units, while four other features collectively have a negligible effect on the prediction.

4.5.1.5 SHAP Force Plot

It enables us to observe the contribution of features to the model's prediction for a given observation. This makes it ideal for explaining to someone how the suggested model came to the conclusion that it did for a particular observation. [43]

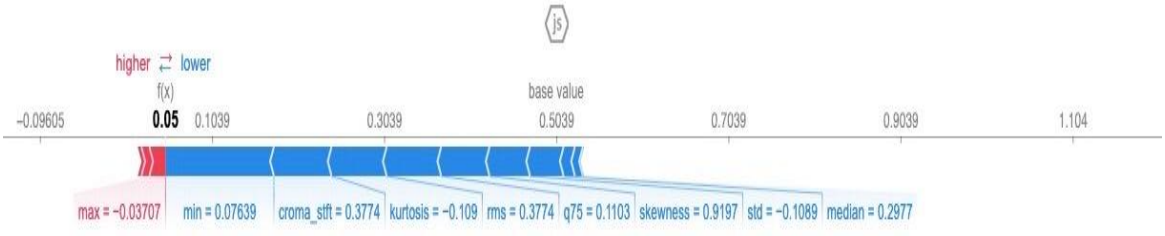

Figure 4.20: Force Plot

The binary target is to find the contribution of each feature to the classification of CVDs. In Figure 4.20, the bold **0.05** is the model score. The model predicts 1 when the scores are higher and 0 when the scores are lower. The elements that were crucial in generating the forecast for the suggested study are displayed in red and blue, where red stands for features that increased the model's score and blue for those that decreased it. Features that had more of an impact on the score are located closer to the dividing boundary between red and blue, and the size of that impact is represented by the size of the bar.

The above simulation results underscore the effectiveness of the RF classifier in heart sound classification, outperforming KNN and SVM. The integration of SHAP values further enhances the interpretability of the model, making it more valuable in a clinical context. These findings contribute to the broader understanding of ML applications in CVD diagnostics.

4.5.2 LIME Analysis

Local Interpretable model-agnostic explanations which is abbreviated as LIME. Unlike attending to the global explanation for the entire population or sample, LIME offers a localized explanation of the model’s prediction for a given instance [24]. Feature importance achieves the goal of explaining how specific features affect individual prediction, which in turn, helps to better understand the model’s behavior on an instance-by-instance basis. LIME was chosen for its ability to provide localized explanations for individual predictions, making it particularly useful for understanding complex models in the context of clinical data. The LIME feature importance bar plot and other visualizations help identify which features most significantly impact a specific prediction, allowing practitioners to interpret the model’s behavior on a case-by-case basis. This level of detail is crucial for validating model decisions, as it helps clinicians understand the rationale behind predictions and enhances trust in the model's outputs.

Figure 4.21 through Figure 4.26 show the LIME feature importance bar plot, feature importance, feature weights plot, cumulative feature importance plot and, feature importance heatmap.

4.5.2.1 Feature Importance Bar Plot

The feature importance bar plot is a visualization created by LIME that illustrates the contributions of the features for a particular prediction.

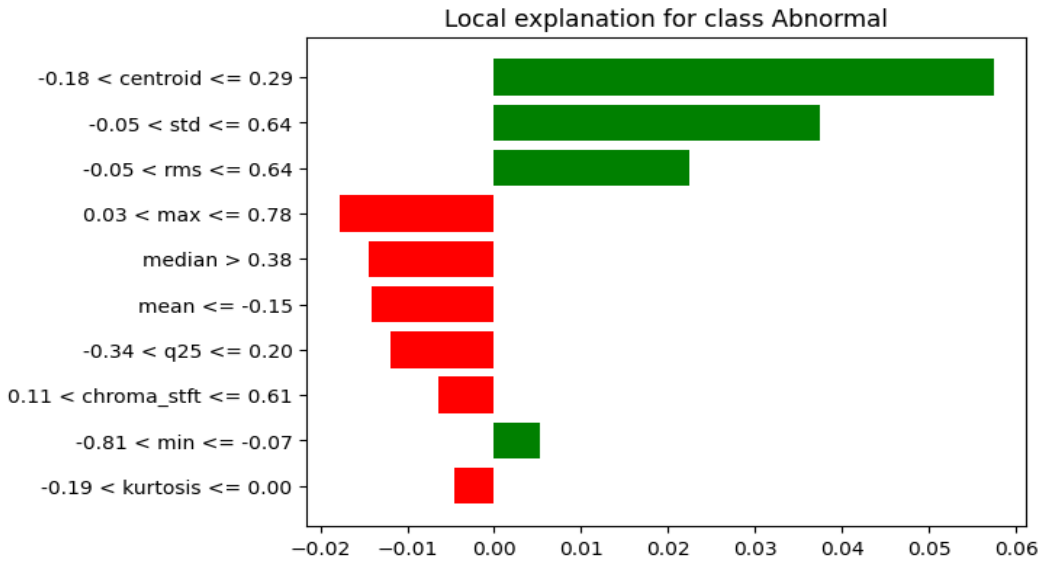

Figure 4.21: Local Explanation for Class Abnormal

The LIME bar plot in Figure 4.21 gives a local interpretation to the specific prediction that has been classified as Abnormal. X-axis shows the partial contribution or the influence of the features towards the predictions made with the magnitude as well as sign of the features. This is manifested in that a result having a green bar pushes the prediction towards the Abnormal class, while a result having a red bar will push the result towards the Normal class. Thus, the features are ranked along the Y-axis according to the obtained contribution values where the features with the highest contribution value are placed in the upper part of the ranking. For instance, the centroid feature (-0.18<centroid≤0.29) has a maximum positive contribution of approximately 0.057. So it has a powerful impact on the model's decision to classify the instance as Abnormal. On the other hand, the maximum value feature (0.03< max≤ 0.78) holds the overall maximum negative attribute contribution of roughly negative 0.02 shows an inclination toward a Normal class.

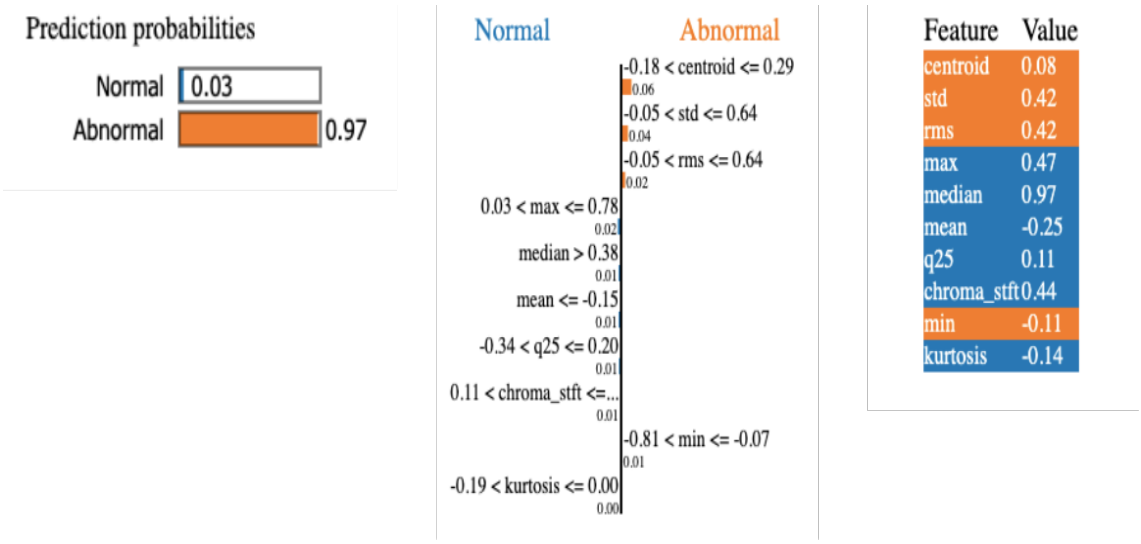

Figure 4.22: Prediction Probabilities

Above Figure 4.22 contains three main pieces of information from left to right: a description of the model results, the values of contribution, and the actual value of each of the characteristics in the model. For the instance being Abnormal, the element has a score of 0.97 and, the score is 0.03 for being Normal. The features increasing the chances of the prototype to be classified as Abnormal are shown in orange on the right, while the features that decrease this probability and move the prototype closer to Normal, are shown in blue on the left. For instance, range of centroid feature of Abnormal is (-0. 18≤ centroid ≤0. 29), and the standard deviation of Abnormal is (-0. 05 < std ≤ 0. 64) has shown the values of potential outcome for Abnormal in positive correlation, while the maximum value of Abnormal is (0. 03 ≤ max ≤ 0. 78) and Median has proven in negative correlation.

4.5.2.2 Local Model Explanations-Feature importance

The Figure 4.23 represents the feature importance in regard to a particular instance. The x-axis present the importance values of each feature, which show how relevant each feature is to the given instance and the model's prediction in particular. The features are listed along y-axis in the model. Regarding the importance of features the 'centroid' has the

highest importance value of approximately 0.06, making it, therefore, closely associated with the function of providing the most crucial input in the model's decision making for this type of instance. 'rms' and 'std' are also considered to be quite vital, with importance levels estimated at 0.04 and slightly less than 0.04, respectively. Features including 'skewness', 'q75', 'kurtosis', as well as 'chroma_stft' have comparatively smaller importance values meaning that they play a smaller role in the prediction of the model. Next, the importance of the features 'min' and 'max' is very low, which indicates that they have almost minimal influence on the decision.

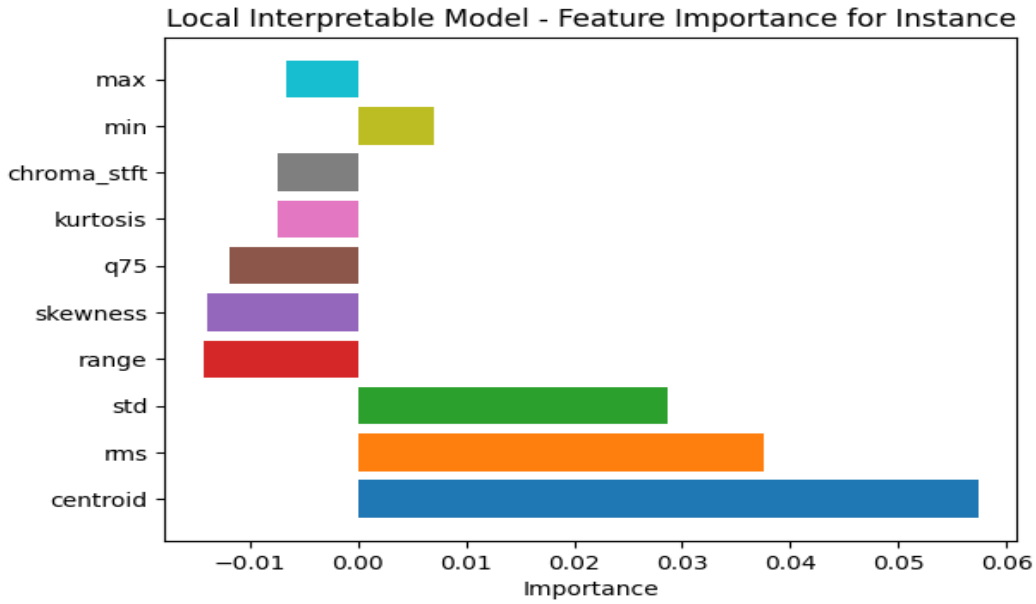

Figure 4.23: Feature Importance for Instance

4.5.2.3 Feature Weights Plot

The feature importance plot shown in Figure 4.24 highlights information on which features are used to arrive at a certain decision by the classifier within the local model explanation. The x-axis of the plot represents feature weights, which denote the contribution made by each of the used features toward the model's decision. A positive value of weight depicts that the feature has a positive contribution and hence it increases the prediction and the 'y' axis illustrates the features of the instance being explained along with their

corresponding feature values. The length of the individual bars represents the contribution of the identified feature to the model's outcome. When feature bars are longer, it implies that the affect or impact of such feature either positive or negative is stronger in making the decision.

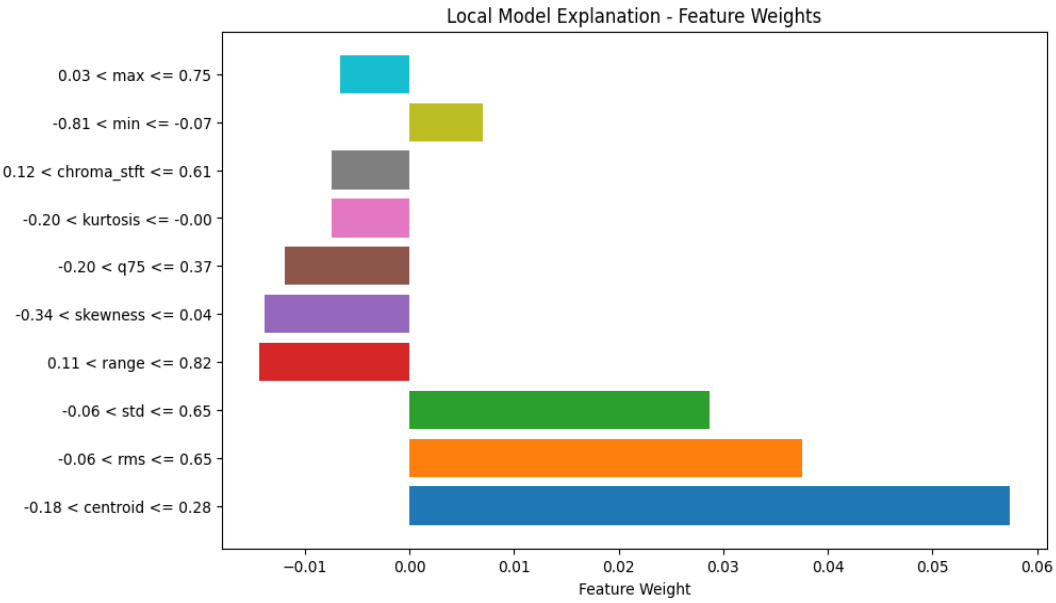

Figure 4.24: Feature Weights

The plot shows that the 'centroid' feature actually has a weight of about 0.06 increases the model prediction. Next to this, the 'rms' and 'std' features also substantially contribute positive values. The 'range' feature on the other hand has a positive contribution but to a lower magnitude. Other aspects locate the bar at close to zero or slightly negative, which are 'skewness', 'q75', 'kurtosis', 'chroma_stft' and 'min'. Significantly, the 'max' feature has a slightly negative effect on the prediction.

4.5.2.4 Cumulative Feature Importance Plot

It allows the decision maker to get an idea of which parameters are most significant and how they collectively affect the model's decision.

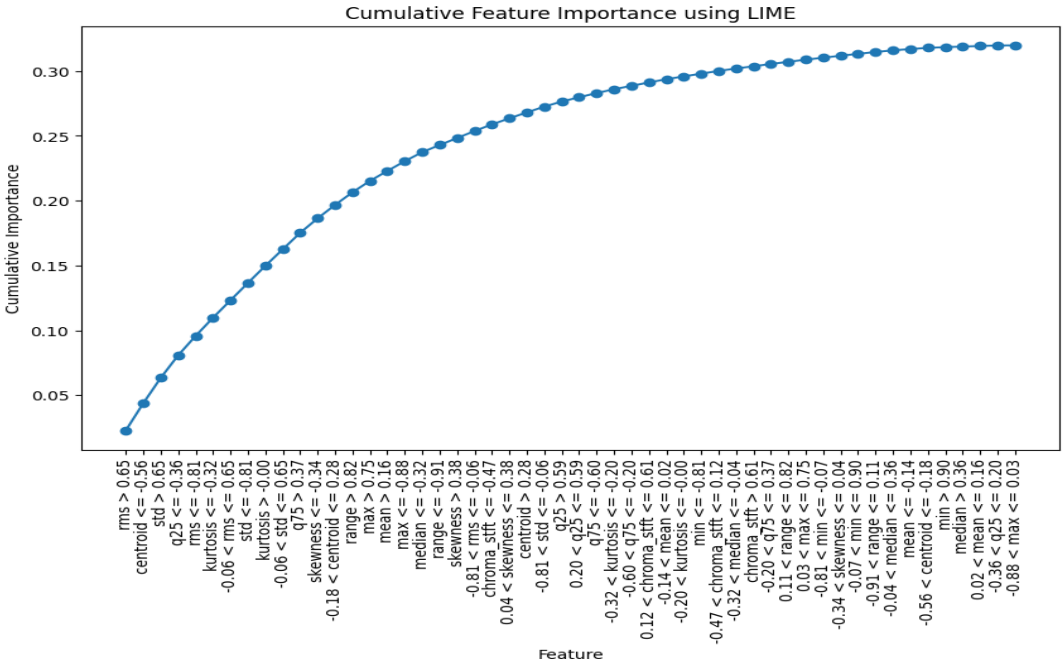

Figure 4.25: Cumulative Feature Importance(LIME)

Understanding of each feature contribution to the model's overall decision is shown in the cumulative feature importance plot in Figure 4.25, calculated by the LIME technique. On the horizontal axis there are listed features, and on the vertical axis the cumulative importance, which sums up the feature's contribution one by one. It continues to rise steep which stands for as the number of important features increase, where the first few and most sensitive are 'centroid', 'rms', and 'std' as they significantly affect the model. With subsequent features the graph begins to flatten which implies that the features such as 'q75', 'chroma_stft', 'max' do not carry as much importance and have less influence to the prediction.

4.5.2.5 Feature Importance Heatmap

The heatmap can be used to summarize the level of contribution that each feature has made to the model's decision-making across different instances of a set and can be used to give insights of the model's behavior.

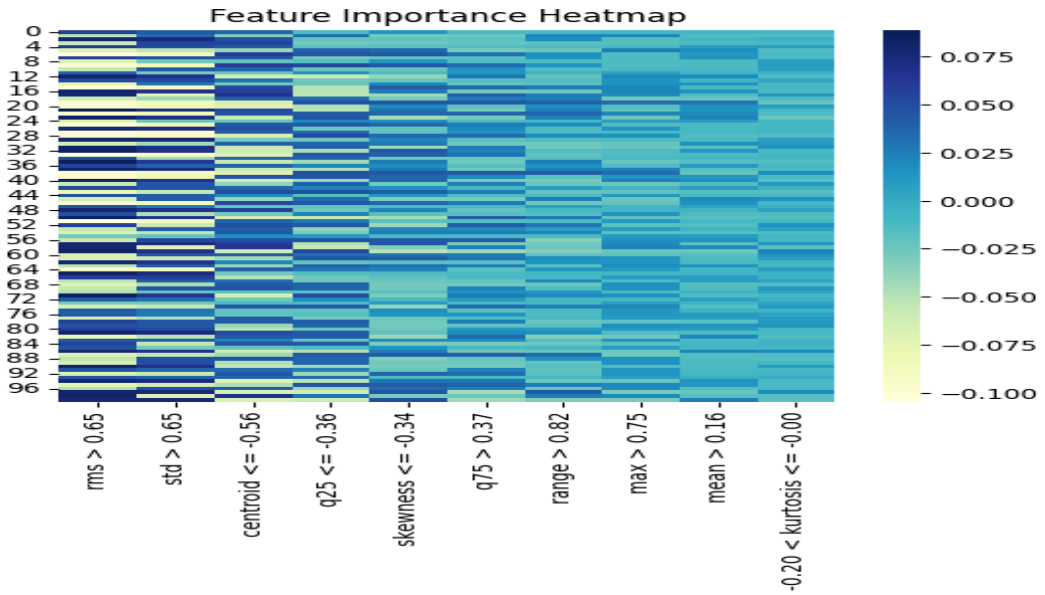

Figure 4.26: Feature Importance Heatmap

The heatmap given in Figure 4.26 represents the feature importance that signifies the concerning features affect the model for different samples in a influencing manner in the dataset. The horizontal line of labels indicate the name of the features, such as ‘rms’, ‘std’, ‘centroid’, other, while the vertical line of labels represents the samples ranging from 0 to 96. Thus, each cell in the heatmap reflects the feature importance of a specific sample and colors range from dark blue to the yellow color. A dark blue color means strong negative influence, light yellow means strong positive influence whereas the middle shade means intermediate influence. For example, the feature rms > 0.65 regularly exhibits light yellow cells, meaning that it has a very positive effect on the predictions of numerous samples. On the other hand, other feature such as skewness <= -0.34 may paint the cells with dark blue showing how the item had a negative influence to some samples on the predictions.

4.6 Summary

This chapter described and discussed the simulation for the classification of heart sounds using the PhysioNet 2016 dataset. Based on the findings, the performance of the Random Forest (RF) classifier demonstrated exceptional accuracy and reliability in classifying heart sounds, with metrics such as a 93.82% accuracy and an F1 score of 93.64%. The detailed analysis of the RF classifier using confusion matrices provided a clear understanding of its predictive capabilities. Additionally, the use of SHAP and LIME enhanced model interpretability by revealing how individual features contribute to predictions. These tools allowed for a comprehensive exploration of feature importance, ultimately supporting better clinical decision-making through enhanced transparency and understanding of the model's behavior.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

This chapter integrates the conclusions of the study concerning the application of XAI to classify CVD employing unsegmented PCG signals. It elaborates on the employment and importance of the SHAP and LIME in the improvement of the interpretability and performance of the specific RF classifier model. Besides, the chapter also qualities a discussion of the limitations of this research and the identified avenues for future research.

5.1 Conclusion

When adopting machine learning in clinical applications, correct classification and interpretability are standard, although because of the lack of large datasets in similar categories, efficient use of resources is compulsory. Thus, this work demonstrates the potential of XAI techniques as a promising direction in solving the problems of classifier efficiency and interpretability. This implies that the current study is concerned with model interpretability and the design of the features for extracting the XAI of CVD classification in patients using unsegmented PCGs.

As part of attempting an efficient classification, a comparative analysis of Random Forest (RF), Support Vector Machine (SVM), and K Nearest Neighbors (KNN) methods was computed. Regarding the problem of feature extraction and model assessment, the introduced

problem was explained, and statistical features were extracted for classification. The RF model interpretability was then improved using SHAP and LIME. Simulation results demonstrate that RF has better performance over SVM and KNN classifiers to demonstrate and enhance the organizational performance, thus providing a better increase in the classification accuracy, precision, recall, specificity, and F1 score.

Finally, the study results reveal that SHAP and LIME interpretability assessment enables the researcher to understand the RF model's decision-making mechanisms, and, therefore, the predictions made can be explained. Therefore, it is concluded that the integration of SHAP and LIME with RF can provide enhanced performance with increased interpretability in the classification of CVDs adopting unsegmented PCGs concerning other classifiers lacking explainability.

5.2 Contributions and Significance

As a result, each of the studies conducted earlier has a gap in which the application of XAI techniques in the classification of CVD has not been given due attention. Therefore, the application of XAI, especially SHAP and LIME in improving the trustworthiness and accuracy of ML classifiers such as Random Forest in the PhysioNet 2016 could be deemed as a relevant and important area of study in the present-day research agenda. CVD classification is one of the most important problems because it describes a large amount of information, which can stably affect the situation with patients and clinical decisions. Thus, this research could be useful in the categorization of CVDs.

5.3 Limitations and Scope of Future Work

1. This research formulated the classification problem for the first time in this manner, focusing on the PhysioNet 2016 dataset. For the sake of simplicity, only statistical features were extracted and analyzed. Future work can explore the impact of incorporating additional types of features, such as frequency domain features, to increase the model's performance.
2. The scalability of the ML classifiers and the SHAP-based interpretability approach needs to be further investigated in larger datasets with higher dimensions. As the number of features increases, the computational complexity of training the classifier and generating SHAP values may become a bottleneck in terms of computational resources and processing time.
3. While this research focused on a specific type of CVDs, future research should investigate the applicability of the proposed methods to other types of biomedical signals with different characteristics.
4. In this research, the focus was on classification performance and interpretability. In future research, we can explore hybrid approaches for the classification of CVDs. Future work could involve integrating traditional ML methods with DL techniques to leverage the strengths of each approach.
5. In the future we can apply the same research framework to combined datasets, such as PhysioNet 2016 and PASCAL, to potentially enhance the model's effectiveness and generalizability in the classification of CVDs.

REFERENCES

[1] P. T. Krishnan, P. Balasubramanian, and S. Umapathy, "Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network," *Phys Eng Sci Med*, vol. 43, no. 2, pp. 505–515, Jun. 2020, doi: 10.1007/S13246-020-00851-W/METRICS.

[2] X. Sun, Y. Yin, Q. Yang, and T. Huo, "Artificial intelligence in cardiovascular diseases: diagnostic and therapeutic perspectives," *Eur J Med Res*, vol. 28, no. 1, pp. 1–11, Dec. 2023, doi: 10.1186/S40001-023-01065-Y/FIGURES/3.

[3] "Cardiovascular diseases (CVDs)." Accessed: Jul. 18, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[4] "Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016 | IEEE Conference Publication | IEEE Xplore." Accessed:May14,2024.[Online].Available:https://ieeexplore.ieee.org/abstract/document/7868816

[5] A. Raghu, D. Praveen, D. Peiris, L. Tarassenko, and G. Clifford, "Engineering a mobile health tool for resource-poor settings to assess and manage cardiovascular disease risk: SMARThealth study," *BMC Med Inform Decis Mak*, vol. 15, no. 1, pp. 1–15, Apr. 2015, doi: 10.1186/S12911-015-0148-4/TABLES/4.

[6] D. Springer, "Mobile phone-based rheumatic heart disease detection," 2015.

[7] "Global Effect of Modifiable Risk Factors on Cardiovascular Disease and Mortality," *New England Journal of Medicine*, vol. 389, no. 14, pp. 1273–1285, Oct. 2023, doi: 10.1056/NEJMOA2206916/SUPPL_FILE/NEJMOA2206916_DISCLOSURES.PDF.

[8] A. Mohd Noor and M. Faiz Shadi, "THE HEART AUSCULTATION. FROM SOUND TO GRAPHICAL."

[9] S. Shu, J. Ren, and J. Song, "Clinical application of machine learning-based artificial intelligence in the diagnosis, prediction, and classification of cardiovascular diseases," Aug. 25, 2021, *Japanese Circulation Society*. doi: 10.1253/circj.CJ-20-1121.

[10] S. A. Singh and S. Majumder, “Short unsegmented PCG classification based on ensemble classifier,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 28, no. 2, pp. 875–889, Jan. 2020, doi: 10.3906/elk-1905-165.

[11] F. Doshi-Velez and B. Kim, “Towards A Rigorous Science of Interpretable Machine Learning”.

[12] G. Vilone and L. Longo, “Notions of explainability and evaluation approaches for explainable artificial intelligence,” *Information Fusion*, vol. 76, pp. 89–106, Dec. 2021, doi: 10.1016/J.INFFUS.2021.05.009.

[13] M. Y. Kim *et al.*, “A Multi-Component Framework for the Analysis and Design of Explainable Artificial Intelligence,” *Mach Learn Knowl Extr*, vol. 3, no. 4, pp. 900–921, Dec. 2021, doi: 10.3390/make3040045.

[14] “Explainable Artificial Intelligence(XAI) - GeeksforGeeks.” Accessed: Jul. 18, 2024. [Online].Available:https://www.geeksforgeeks.org/explainableartificialintelligencexai/

[15] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, and C. Fookes, “A Robust Interpretable Deep Learning Classifier for Heart Anomaly Detection without Segmentation,” *IEEE J Biomed Health Inform*, vol. 25, no. 6, pp. 2162–2171, Jun. 2021, doi: 10.1109/JBHI.2020.3027910.

[16] H. S. Salem, M. Y. Pudza, and Y. Yihdego, “Harnessing the energy transition from total dependence on fossil to renewable energy in the Arabian Gulf region, considering population, climate change impacts, ecological and carbon footprints, and United Nations’ Sustainable Development Goals,” *Sustainable Earth Reviews*, vol. 6, no. 1, Sep. 2023, doi: 10.1186/S42055-023-00057-4.

[17] J. Ralston, K. S. Reddy, V. Fuster, and J. Narula, “Cardiovascular Diseases on the Global Agenda: The United Nations High Level Meeting, Sustainable Development Goals, and the Way Forward,” Dec. 01, 2016, *Elsevier B.V.* doi: 10.1016/j.gheart.2016.10.029.

[18] “1: The anatomy of the heart. | Download Scientific Diagram.” Accessed: Jul. 30, 2024. [Online]. Available: https://www.researchgate.net/figure/The-anatomy-of-the-heart_fig1_228970175

[19] V. N. Varghees and K. I. Ramachandran, “A novel heart sound activity detection framework for automated heart sound analysis,” *Biomed Signal Process Control*, vol. 13, no. 1, pp. 174–188, 2014, doi: 10.1016/j.bspc.2014.05.002.

[20] S. Chandra, M. Krishanthi, P. J. Octavian, and A. Postolache, “Smart Sensors, Measurement and Instrumentation 29 Modern Sensing Technologies.” [Online]. Available: http://www.springer.com/series/10617

[21] S. Behbahani, “A hybrid algorithm for heart sounds segmentation based on phonocardiogram,” *J Med Eng Technol*, vol. 43, no. 6, pp. 363–377, Aug. 2019, doi: 10.1080/03091902.2019.1676321.

[22] D. Lloyd-Jones *et al.*, “Executive summary: Heart disease and stroke statistics-2010 update: A report from the american heart association,” Feb. 2010. doi: 10.1161/CIRCULATIONAHA.109.192667.

[23] “Stethoscope/Auscultation landmarks : 네이버 블로그.” Accessed: Jul. 18, 2024. [Online]. Available: https://m.blog.naver.com/generalfit/221031965778

[24] “Explainable AI, LIME & SHAP for Model Interpretability | Unlocking AI’s Decision-Making | DataCamp.” Accessed: Jul. 18, 2024. [Online]. Available: https://www.datacamp.com/tutorial/explainable-ai-understanding-and-trusting-machine-learning-models

[25] Y. ; Zhang *et al.*, “Applications of Explainable Artificial Intelligence in Diagnosis and Surgery,” *Diagnostics 2022, Vol. 12, Page 237*, vol. 12, no. 2, p. 237, Jan. 2022, doi: 10.3390/DIAGNOSTICS12020237.

[26] M. Guven and F. Uysal, “A New Method for Heart Disease Detection: Long Short-Term Feature Extraction from Heart Sound Data,” *Sensors 2023, Vol. 23, Page 5835*, vol. 23, no. 13, p. 5835, Jun. 2023, doi: 10.3390/S23135835.

[27] J. Li *et al.*, “Heart Sound Signal Classification Algorithm: A Combination of Wavelet Scattering Transform and Twin Support Vector Machine”, doi: 10.1109/ACCESS.2019.2959081.

[28] S. A. Singh and S. Majumder, “CLASSIFICATION OF UNSEGMENTED HEART SOUND RECORDING USING KNN CLASSIFIER,” *J Mech Med Biol*, vol. 19, no. 4, Jun. 2019, doi: 10.1142/S0219519419500258.

[29] T. H. Chowdhury, K. N. Poudel, and Y. Hu, “Time-Frequency Analysis, Denoising, Compression, Segmentation, and Classification of PCG Signals,” *IEEE Access*, vol. 8, pp. 160882–160890, 2020, doi: 10.1109/ACCESS.2020.3020806.

[30] Y. R. Chien, K. C. Hsu, and H. W. Tsao, “Phonocardiography Signals Compression with Deep Convolutional Autoencoder for Telecare Applications,” *Applied Sciences*

2020, Vol. 10, Page 5842, vol. 10, no. 17, p. 5842, Aug. 2020, doi: 10.3390/APP10175842.

[31] K. N. Khan *et al.*, "Deep learning based classification of unsegmented phonocardiogram spectrograms leveraging transfer learning," *Physiol Meas*, vol. 42, no. 9, p. 095003, Sep. 2021, doi: 10.1088/1361-6579/AC1D59.

[32] S. K. Ghosh, R. N. Ponnalagu, R. K. Tripathy, G. Panda, and R. B. Pachori, "Automated Heart Sound Activity Detection From PCG Signal Using Time-Frequency-Domain Deep Neural Network," *IEEE Trans Instrum Meas*, vol. 71, 2022, doi: 10.1109/TIM.2022.3192257.

[33] F. Demir, A. Şengür, V. Bajaj, and K. Polat, "Towards the classification of heart sounds based on convolutional deep neural network," *Health Inf Sci Syst*, vol. 7, no. 1, pp. 1–9, Dec. 2019, doi: 10.1007/S13755-019-0078-0/METRICS.

[34] F. A. Khan, A. Abid, A. Abid, M. S. Khan, and M. S. Khan, "Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features," *Physiol Meas*, vol. 41, no. 5, p. 055006, Jun. 2020, doi: 10.1088/1361-6579/AB8770.

[35] S. A. Singh, S. Majumder, and M. Mishra, "Classification of short unsegmented heart sound based on deep learning," *I2MTC 2019 - 2019 IEEE International Instrumentation and Measurement Technology Conference, Proceedings*, vol. 2019-May, May 2019, doi: 10.1109/I2MTC.2019.8826991.

[36] S. Perri, R. De Fazio, L. Spongano, M. De Vittorio, L. Patrono, and P. Visconti, "Machine Learning Algorithms for Processing and Classifying Unsegmented Phonocardiographic Signals: An Efficient Edge Computing Solution Suitable for Wearable Devices," *Sensors 2024, Vol. 24, Page 3853*, vol. 24, no. 12, p. 3853, Jun. 2024, doi: 10.3390/S24123853.

[37] M. Chowdhury, C. Li, and K. Poudel, "Combining Deep Learning with Traditional Machine Learning to Improve Phonocardiography Classification Accuracy," in *2021 IEEE Signal Processing in Medicine and Biology Symposium, SPMB 2021 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/SPMB52430.2021.9672296.

[38] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, H. Ghaemmaghami, and C. Fookes, "Understanding the Importance of Heart Sound Segmentation for Heart Anomaly Detection".

[39] “Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016 | IEEE Conference Publication | IEEE Xplore.” Accessed:May14,2024.[Online].Available:https://ieeexplore.ieee.org/abstract/document/7868816

[40] A. K. Mittra and N. K. Choudhari, “Time-frequency analysis of foetal heart sound signal for the prediction of prenatal anomalies,” *J Med Eng Technol*, vol. 33, no. 4, pp. 296–302, May 2009, doi: 10.1080/03091900802454384.

[41] A. Ogunpola, F. Saeed, S. Basurra, A. M. Albarrak, and S. N. Qasem, “Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases,” *Diagnostics*, vol. 14, no. 2, Jan. 2024, doi: 10.3390/diagnostics14020144.

[42] “Explainable AI, LIME & SHAP for Model Interpretability | Unlocking AI’s Decision-Making|DataCamp.”Accessed:Jul.18,2024.[Online].Available:https://www.datacamp.com/tutorial/explainable-ai-understanding-and-trusting-machine-learning-models

[43] “SHAP Force Plots for Classification | by Max Steele (they/them) | Medium.” Accessed: Jul. 18, 2024. [Online]. Available: https://maxsteele731.medium.com/shap-force-plots-for-classification-d30be430e195