Independent University **Bangladesh (IUB)**

IUB Academic Repository

Graduate Thesis 2025

2025-12

Coronary Microvascular Disease Detection Using Ensemble Learning and Advanced ECG Signal Processing

Tabassum, Yousra

IUB

http://ar.iub.edu.bd/handle/11348/1034

Downloaded from IUB Academic Repository



Coronary Microvascular Disease Detection Using Ensemble Learning and Advanced ECG Signal Processing

December 2025

Yousra Tabassum
ID: 2222229

Anika Islam
ID: 2220151

Nusrat Sultana
ID: 2221125

Department of Computer Science and Engineering
Independent University, Bangladesh

Supervised by

Ashraful Islam, PhD
Assistant Professor
Department of Computer Science and Engineering
Independent University, Bangladesh

Attestation

We are aware of the fact that plagiarism is strictly prohibited and it conflicts with our university's rules and regulations. We guarantee the authenticity of our work. We have used some copyrighted material and models of others in our project which has been properly cited following the international standards proper guidelines.

Author Name: Yousra Tabassum
Signature: _____

Author Name: Anika Islam
Signature: _____

Author Name: Nusrat Sultana
Signature: _____

Evaluation Committee

Supervision Panel

.....	.....
Dr. Ashraful Islam	Dr. Rashedur Rahman
Supervisor	Co-supervisor

Examiners

.....	.....
Dr. AKM Mahbubur Rahman (Internal)	Mr. Sanzar Adnan Alam (Internal)

.....

Mr. Asif Md. Akhlak (External Examiner)

Office Use

.....	.....
Dr. AKM Mahbubur Rahaman	Dr. Saadia Binte Alam
Program Coordinator	Head of the Department

.....

Prof. Dr. M. Arshad Momen

Director, Office of the Graduate Studies, Research and Industry Relations

.....

Library

Acknowledgement

We would like to express our sincere gratitude to our respected supervisor, Ashraful Islam, PhD, Assistant Professor, Independent University, Bangladesh. His dedicated guidance, insightful suggestions, remarkable patience, and the generous investment of his valuable time played a pivotal role in the successful completion of our project. Our supervisor's expertise and mentorship greatly enhanced the quality and depth of our work, and we are truly appreciative of his unwavering support throughout our academic journey.

Additionally, our heartfelt gratitude goes to Center for Computational & Data Sciences (CCDS) for their extraordinary help and consistent support. The collaborative and resourceful environment provided by CCDS significantly contributed to the development and execution of our project. Moreover, the encouragement and assistance have been instrumental in shaping our academic and research fields. We are grateful for the opportunities and resources extended to us during our research, which undoubtedly enriched our learning experience.

We acknowledge and appreciate the contributions of all individuals who supported us, both directly and indirectly, during this endeavor.

Abstract

Coronary Microvascular Disease (CMD) is a type of ischemic heart disease which affects the very small blood vessels of the heart. It causes chest pain and reduces blood flow. As the large arteries look normal in standard test, CMD is hard to detect. Heart diseases are usually checked using electrocardiograms (ECG), but small changes related to CMD are often too minimal to notice it. This research presents if machine learning can automatically identify CMD from 12 lead ECG signals taken during an Adenosine stress test.

The study used a dataset of 97 subjects, including 19 participants with CMD and 78 healthy individuals, derived from ECG data. The ECG was recorded into three stages, Rest, Stress, and Recovery. During the stress phase, a lot of noise was present because of heavy breathing. To solve this, Dual Path preprocessing pipeline was used to clean the noise and measure the features correctly. From the ECG signals, the study extracted 612 features related to heart rate and Morphological features. Using mutual information and correlation filtering, it was reduced to the top 12 most critical biomarkers. An ensemble model, combination of three models Support Vector Machine, Random Forest, and XGBoost was trained and validated using a Leave-One-Subject-Out (LOSO) cross-validation strategy.

ECG cleaning methods are insufficient and missed many CMD cases, but the custom preprocessing that was used in this study improved signal quality. The feature selection analysis represents the physiological changes. The final Ensemble model achieved a high classification performance with an F1-Score of 0.69, a Recall (Sensitivity) of 68%, Precision 0.69 and an Accuracy of 82%.

This study shows that CMD can be detected using ECG signals. The noisy data can be cleaned using correct methods. By combining changes in the morphological ischemia markers with heart rate patterns, accurate CMD identification is possible. This approach can be affordable and easy to use and could improve early diagnosis of CMD.

Contents

Attestation i

Acknowledgement iv

Abstract v

1 Introduction 1

1.1 General Background About the Disease 1
1.2 The Specific Dilemma 2
1.3 Why is this Still Unsolved 2
1.4 Our Idea 3
1.5 Hypothesis 4
1.6 CMD in the Perspective of Bangladesh 5
1.7 UN Sustainable Development Goals 5
1.8 Research Objectives and Research Questions 6
1.8.1 Objectives 6
1.8.2 Research Questions 6
1.9 Report Outline 7

2 Literature Review 10

2.1 Literature Review of Papers 10
2.1.1 Method 15
2.1.2 Search Strategy 15
2.1.3 Data Extraction 18
2.1.4 Literature Search Outcomes 20
2.1.5 Findings 22

3 Methodology 24

3.1 Dataset Description 24
3.2 Data Pre-processing 25
3.3 Feature Selection 27
3.4 Model Development 28

3.4.1 Logistic Regression 28
3.4.2 Support Vector Machine (Kernel - RBF) 29
3.4.3 Random Forest 30
3.4.4 K - Nearest Neighbor (KNN) 30
3.4.5 XGBoost 31
3.4.6 MLP - MultiLayer Perceptron 31
3.4.7 Ensemble Learning 32
3.5 Evaluation Metrics 32
3.5.1 Cross Validation 34

4 Results & Analysis **35**
4.1 Logistic Regression 35
4.2 Support Vector Machine (Kernel - RBF) 36
4.3 K-Nearest Neighbor (KNN) 37
4.4 Random Forest 38
4.5 XGBoost 39
4.6 MLP 40
4.7 Ensemble (XGB + RF + SVM) 41

5 Discussion **43**
5.1 Objective 43
5.2 The Critical Role of Preprocessing 43
5.3 Feature Importance and Physiological Markers 44
5.3.1 The Value of “Delta” Features 44
5.3.2 Morphology vs. HRV 45
5.4 Model Performance and Class Imbalance 45
5.4.1 Failure of Simple Models 45
5.4.2 Success of the Ensemble Model 45
5.5 Limitations of the Study 46

6 Conclusion and Future Work **47**
6.1 Summary of the Study 47
6.2 Key Conclusions 47
6.3 Contributions to the Field 48
6.4 Future Work 48
6.4.1 Improving Explainability and Clinical Trust 48
6.4.2 Expanding to Deep Learning with Larger Datasets 49
6.4.3 External and Multi-Centre Validation 49
6.4.4 Real-Time Implementation 49
6.5 Final Remarks 49

Appendix A: Abbreviations 50

Bibliography **52**

List of Figures

2.1 PRISMA flow diagram of the literature search and selection process. . . . 17
2.2 Data Extraction Themes 19
2.3 Word cloud based on the keywords. 20

3.1 Overview of the dataset. 25

4.1 Confusion Matrix for Logistic Regression 36
4.2 Confusion Matrix for SVM (RBF Kernel) 37
4.3 Confusion Matrix for KNN 38
4.4 Confusion Matrix for Random Forest 39
4.5 Confusion Matrix for XGBoost 40
4.6 Confusion Matrix for MLP 41
4.7 Confusion Matrix for Ensemble model 42

List of Tables

2.1 Research Questions for SLR and Their Rationale 12
2.2 Summary of Machine Learning Studies for Cardiovascular and Coronary Microvascular Disease Diagnosis 13
2.3 Lists of Inclusion and Exclusion Criteria 16
2.4 Summary of Common Limitations Reported in CMD ML/DL Studies . . 21

4.1 Performance Comparison of Different Machine Learning Models for CMD Classification 42

Chapter 1

Introduction

1.1 General Background About the Disease

Heart disease is a broad term for a range of conditions which affect the heart or blood vessels, with coronary artery disease being the most common. It includes different problems; among them, Cardiovascular disease remains the leading cause of death worldwide. It has a great impact on the health systems and societies. In the United States, coronary heart disease has accounted for hundreds of thousands of deaths in recent years [1]. According to the Centers for Disease Control and Prevention (CDC), approximately 1 in 5 deaths is caused by heart disease in the United States [2].

Coronary Microvascular Disease (CMD) is a type of Ischemic heart disease, which is also called coronary microvascular dysfunction, microvascular angina, or ischemia with non-obstructive coronary arteries (INOCA). CMD affects the small coronary arterioles and capillaries, instead of the large epicardial vessels. When the coronary microvessels fail to dilate appropriately or constrict inappropriately, it disrupts blood flow, which can produce ischemic symptoms despite no significant epicardial stenosis [3].

CMD is prevalent among women. Previous research indicates that women who are facing angina and non-obstructive coronary arteries often have microvascular dysfunction [4]. CorMicA has shown that invasive coronary function testing can be used to identify CMD, which can guide stratified medical therapy to improve patient-reported outcomes [5].

Accurate diagnosis of CMD is a great challenge because routine coronary angiography assesses epicardial luminal obstruction but not microvascular function. As CMD is almost an underdiagnosed form of ischemic heart disease that contributes to persistent angina, increased risk of cardiovascular outcomes, early identification and management of CMD is crucial now. Current diagnostic methods, such as Positron Emission Tomography (PET), are expensive and not widely accessible [6][7].

1.2 The Specific Dilemma

CMD is a complex cardiovascular condition which indicates dysfunction in the small vessels of the heart. It can lead to myocardial ischemia without the presence of obstructive coronary artery disease [8]. CMD remains underdiagnosed and poorly understood in clinical practice, and this is largely due to the limitations of current diagnostic approaches [9]. This disease can only be diagnosed by invasive, expensive, or technically demanding procedures such as coronary flow reserve (CFR) measurement or intravascular imaging [10]. These limitations highlight the necessity for widespread, early, and non-invasive detection as well as reliable risk stratification of CMD patients [11].
Meanwhile, CMD is known to be associated with adverse cardiac events, but there is still no standardized pathway for its classification or early diagnosis using accessible, low-cost, and non-invasive tests [12]. This dilemma highlights an urgent need for innovative diagnostic tools capable of integrating heterogeneous clinical and imaging data to identify CMD with higher accuracy and clinical relevance.
In recent years, many people especially women experienced angina and symptoms of myocardial ischemia but their angiograms showed no obstruction in their large epicardial coronary arteries, this condition is known as Ischemia with Non-Obstructive Coronary Arteries (INOCA) [[13]. In many cases, the real underlying cause is CMD, which involves dysfunction in the small vessels that run within the myocardium and regulates blood flow [14]. CMD affects the small blood vessels of the heart that are too tiny to be seen in normal angiography. When these vessels fail to deliver enough oxygen, it leads to chest pain and heart muscle stress even without major artery blockage [15]. As a result, patients may present with symptoms that strongly match ischemia with chest pain, shortness of breath, but their angiograms appear “normal” or show only mild disease. This situation creates frustration for both patients and clinicians and leads to diagnostic uncertainty and confusion. This became a real burden for the healthcare system. Patients with CMD often suffer from angina, which reduces their quality of life and makes them visit the hospital repeatedly, and worsens cardiovascular outcomes compared with those without microvascular problems [16]. CMD represents a silent but concerning public health challenge that is often overlooked despite its serious impact on patient wellbeing and health [13].

1.3 Why is this Still Unsolved

CMD remains a challenging condition to diagnose and classify, despite its growing recognition in clinical importance. One major obstacle is the reliance on invasive diagnostic procedures, such as Coronary Flow Reserve (CFR), Index of Microcirculatory Resistance (IMR) and Coronary Flow Velocity Reserve (CFVR). These methods give

highly accurate results but they are costly, technically demanding, and carry procedural risks [17], [18]. Because large-scale, high-quality datasets can only be obtained through these invasive techniques, researchers face significant limitations when developing robust non-invasive tools, including machine learning models that require well-labeled data.
The heterogeneity of CMD further complicates diagnosis. The disease is influenced by systemic risk factors, patient demographics, and subtle clinical and functional markers, making its presentation highly variable [19]. Unlike conditions identified by a single biomarker, CMD manifests as a combination of patterns across clinical symptoms, imaging findings, and laboratory results. Integrating these diverse datasets into a coherent diagnostic framework has proven difficult and even advanced machine learning methods often struggle to capture the complex and non-linear interactions that may reveal the underlying dysfunction.
Adding to this challenge, CMD encompasses multiple endotypes with distinct mechanisms. Structural microvascular changes, such as capillary rarefaction, coexist alongside functional abnormalities like endothelial dysfunction and vasospasm, each associated with different risk profiles and therapeutic implications [18], [19]. This phenotypic diversity makes a simple binary classification insufficient, emphasizing the need for diagnostic approaches that can differentiate between disease subtypes and provide clinically meaningful guidance.
Overall, CMD remains an unsolved problem because it requires solutions that overcome the limitations of invasive testing, integrate heterogeneous and multimodal data, and address the complexity and variability inherent in the disease. These challenges highlight the urgent need for innovative computational tools capable of accurate, non-invasive diagnosis and effective patient stratification.

1.4 Our Idea

The main idea of this study is to develop an automated and non-invasive system for the detection of CMD using Electrocardiogram (ECG) signals and machine learning (ML) techniques. CMD is difficult to diagnose with conventional methods because it affects the small blood vessels of the heart and often does not produce clear abnormalities in standard ECG tests. As a result, many patients remain undiagnosed or are diagnosed late.

In this study, we focus on extracting useful physiological information from ECG signals, recorded during both rest and stress conditions. Instead of relying on a single type of feature, we combine two important group of features: morphological features, which describe the shape and changes in ECG waves, and heart rate variability (HRV) features, which reflect the activity of the autonomic nervous system. By combining these two types of information, the model can capture both electrical and regulatory changes in the heart

that are linked to CMD.

A major part of our idea is to design a strong signal preprocessing pipeline. Since ECG signals often contain noise and motion artifacts in all phases of recording, proper cleaning is essential before feature extraction. Through careful filtering, artifact removal, and beat selection, we aim to ensure that only high-quality heartbeats are used for analysis. This improves the reliability of the extracted features and helps the machine learning models learn meaningful patterns [20].

After feature extraction, we apply and compare multiple machine learning models to classify non-CMD subjects and CMD patients. These include traditional classifiers as well as ensemble-based methods. The goal is to identify the model that provides the best balance between accuracy, sensitivity, and overall reliability. Special attention is given to the problem of class imbalance, as in the dataset, CMD cases are way fewer than healthy cases.

Overall, the proposed idea combines advanced ECG preprocessing, meaningful feature extraction from both morphology and HRV, and robust machine learning models to build an effective CMD detection framework. This approach aims to improve early diagnosis using a non-invasive, low-cost, and clinically practical solution.

1.5 Hypothesis

CMD is a complex cardiovascular condition that often remains undetected through traditional diagnostic methods. Recent improvements in ML have demonstrated hopeful potential for identifying such diseases. However, the method for diagnosing CMD is the CFR, which is derived from PET/CT assessment. Since, PET/CT is expensive and time-consuming, also is not available in many clinical institutions, our study highlights the need of an alternative, easily accessible and low in cost option for detecting CMD early. The dataset used in the study consists the complete physiological, biochemical and psycological data of 97 participants. All subjects went through an 18 minute ECG, recorded in three stages, rest, adenosine stress and recovery, collected in 500 Hz, 16-bit precision. The dataset also consists of the CFR value, with the patients labeled 1 if they are CMD positive and 0, if they are CMD negative [21]. But given the fact that how time-consuming and costly PET/CT can be, our study aims to evaluate whether it is possible to distinguish CMD positive or negative using only the ECG derived features. Our idea is to use ML models to derive important features such as morphological and HRV based features from the ECG data for each patients and classify CMD or non CMD patients from that. The purpose of the study is also to determine if ECG alone has sufficient diagnostic information to predict CMD without relying on costly imaging techniques.

Null Hypothesis (H_0): ML models trained only on ECG-derived features do not achieve significant accuracy in classifying CMD, compared to other costly and more direct

medical imaging techniques, indicating that ECG alone is not sufficient for precise and reliable CMD diagnosis.

Alternative Hypothesis (H_1): ML models trained only on ECG-derived features achieve significantly better accuracy in classifying CMD, indicating that ECG signals contain useful diagnostic information for CMD detection.

Hypothesis Testing Plan: To test these hypothesis, several ML models such as Logistic Regression, Random Forest, KNN, SVM, XGBoost, and MLP will be trained using ECG features extracted from Rest, Adenosine Stress, and Recovery phases. The accuracy of the results will be evaluated using cross-validation techniques and standard metrics.

Consistent accuracy in classifying CMD using ECG-derived features will result in rejecting the null hypothesis and accepting the alternative hypothesis, supporting the idea of ECG-based CMD prediction as a low-cost and non-invasive method.

1.6 CMD in the Perspective of Bangladesh

CMD is mainly recognized in Bangladesh through hospital and specialist reports instead of population surveys. There are no nationally representative wide studies that quantify CMD specifically. Clinicians describe that CMD is an under diagnosed and significant disease but its actual prevalence is unknown in Bangladesh [22]. When patients face persistent angina and they undergo coronary angiography and no obstructive coronary lesions found at that moment CMD identified [23]. Compared to other disease CMD gets less attention due to limited diagnostic facilities which are only available in a few centers.

1.7 UN Sustainable Development Goals

The UN Sustainable Development Goals (SDG) is a set of 17 goals formed by the United Nations (UN) in 2015 to achieve peace and prosperity for the whole world. They address several challenges and also aim to improve health which is the aim of our study [24].

Out of 17, there are 2 goals in the UN SDG which our study directly contributes to.

1. **Good Health and Well-Being:** This goal is to ensure good health and promote well-being for all at all ages. Our study directly aims to contribute to one of the most severe cardiovascular conditions, CMD. As CMD is often hard to diagnose due to high expanses and time-consuming medical tests such as PET/CT, our study focuses on classifying this issue using ECG based machine learning techniques, which is low in cost and also easily accessible in CMD detection. Early detection through

such easily available methods can reduce risk of severe cardiac complications. Our study helps directly improving the accuracy of CMD diagnosis, which strongly contributes to the third goal of UN SDG, Good Health and Well-Being [24].

2. **Quality Education:** This goal focuses on ensuring inclusive and equitable quality education and promoting lifelong learning opportunities for all. Our study contributes to this goal by combining healthcare and machine learning, allowing students and researchers to gain practical knowledge in medical signal processing and machine learning. By working with real ECG data and machine learning models for CMD detection, this research supports hands-on learning and skill development in biomedical and data-driven healthcare research [24].

1.8 Research Objectives and Research Questions

1.8.1 Objectives

This research aims to develop a machine learning based automated detection of CMD using ECG signals. The primary goal is to enhance CMD screening by designing an effective preprocessing, feature extraction, and classification pipeline using ECG data. The key contributions of this study are:

1. Design and evaluation of a robust ECG preprocessing and feature extraction pipeline to reduce noise and enhance the discriminative power of morphological, HRV, and non-linear ECG features for accurate CMD classification.

2. Systematic comparison of classical machine learning models and deep learning approaches, including ensemble techniques, using subject-level cross-validation to ensure reliable CMD detection.

3. Application of feature selection technique to identify the most significant ECG features and biomarkers contributing to CMD classification and supporting ECG-based CMD screening.

1.8.2 Research Questions

The study is based on four main research questions (RQs) that aim to examine the challenges and opportunities of using ECG signals and machine learning models for classifying CMD.

1. **RQ1:** Can CMD be accurately identified using non-invasive ECG signals, and how do preprocessing steps and feature extraction methods influence classification performance?

- **Rationale:** This question examines whether ECG signals alone provide sufficient information to detect CMD. It also investigates how different preprocessing techniques (noise removal, signal cleaning) and feature extraction approaches (morphological, HRV, entropy-based methods) influence the accuracy and reliability of ML models. Understanding these factors is essential for developing a dependable and effective CMD detection system.

2. **RQ2:** Which ECG features (morphological, HRV, entropy) and phases (Rest, Stress, Recovery) contribute most to CMD detection, and can lead-specific analysis offer additional diagnostic insights?

   - **Rationale:** This question explores the importance of different feature types and ECG phases when identifying CMD. By analyzing feature relevance and examining lead-specific signals, the study aims to determine which aspects of the ECG are most informative, enabling more focused and interpretable CMD classification strategies.

3. **RQ3:** How do small sample sizes and class imbalance in CMD datasets affect the performance of machine learning and deep learning models, and to what extent can subject-level cross-validation strategies such as Leave-One-Subject-Out (LOSO) mitigate these challenges to ensure reliable generalization?

   - **Rationale:** This question addresses dataset-related challenges, such as limited sample availability and uneven CMD vs. control distributions, which may lead to biased ML/DL predictions. It evaluates the effectiveness of validation techniques like LOSO to ensure that model performance reflects real-world applicability and minimizes overfitting to a few subjects.

1.9 Report Outline

Chapter 1 (Introduction):
This chapter presents the background of heart disease specially a particular focus on CMD, which is a type of ischemic heart disease that remains underdiagnosed condition. It explains the main idea behind this study and the impact of CMD. The diagnostic process become a challenge because regular angiogram cannot detect it and other diagnostic test like PET/CT and CFR assessment are costly, invasive and not available in many hospitals. Regular ECG signals cannot detect the small signals but it is one of the most widely used, low cost and non-invasive techniques. Advance ML models can detect the subtle signals from ECG morphology and HRV

features that may reflect microvascular disease. It also highlights the relevance of the study to UN SDGs, identifies the research gaps, shows the objectives and Research Questions (RQs).

Chapter 2 (Systematic Literature Review):
This chapter shows that previous studies applied many ML and DL models to detect heart diseases, including CMD. It also follows the PRISMA guidelines and describes the methodology, search Strategy, inclusion and exclusion criteria, data extraction, literature search outcomes and findings. This chapter includes five SLR specific research questions and addresses the data modalities used, summarizes the methods used in the previous studies, evaluation techniques and results.

Chapter 3 (Methodology):
This chapter presents the methodology to detect CMD from ECG signals. ECG features were extracted from all 12 leads signals and aggregated using mean and standard deviation for each subject. Features were computed separately for rest and delta scores (stress–rest and recovery–stress). The final dataset consists of 94 subjects with 612 features. Feature selection was applied using delta-score analysis and mutual information to keep only informative features. Machine learning models were developed including Logistic Regression, SVM with RBF kernel, Random Forest, KNN, XGBoost, MLP and an ensemble model using soft voting mechanism. Model performance was evaluated using Leave One Subject Out cross-validation and standard evaluation metrics such as accuracy, precision, recall, F1-score and confusion matrix to ensure reliable CMD classification.

Chapter 4 (Results & Analysis):
This chapter presents and analyzes the performance of different machine learning models used for CMD detection. The results of Logistic Regression, SVM (RBF kernel), KNN, Random Forest, XGBoost, MLP and an ensemble model are evaluated using confusion matrices and standard classification metrics such as accuracy, precision, recall and F1-score. Logistic Regression showed weak performance in identifying CMD patients while SVM and MLP achieved moderate results. Random Forest and XGBoost demonstrated stronger and more stable performance in distinguishing CMD from non-CMD subjects. The ensemble model combining XGBoost, Random Forest and SVM achieved the best overall results. This provides the highest accuracy, recall and F1-score.

Chapter 5 (Discussion):
This chapter discusses the overall findings of the research addressing the stated research objectives and research questions. It explains how ECG preprocessing, feature selection, and machine learning models contributed to CMD detection. The chapter also shows the importance of advanced noise reduction techniques, the role

of stress and recovery delta features, and the use of morphological and HRV features. It also describes the model performance under class imbalance, emphasizing the strong performance of ensemble learning and macro-averaged evaluation metrics. Lastly, the limitations of the study are discussed to support the results and future improvements.

Chapter 6 (Conclusion & Future Work):
This chapter summarizes the main contributions and outcomes of the study, demonstrating the practicality of non-invasive CMD detection using ECG signals and machine learning. It highlights key conclusions related to preprocessing strategies, feature importance, and ensemble model performance. The chapter also explains the contributions of the study in biomedical signal processing and clinical decision support systems. Finally, it presents future research directions, including improving model explainability, expanding to larger multi-center datasets, external validation, and real-time clinical implementation.

Appendix:
Appendix A, presents the complete list of all abbreviations used in this report for reference.

Chapter 2

Literature Review

2.1 Literature Review of Papers

Cardiovascular Disease (CVD) and CMD became major health concerns and needed to explore advanced diagnostic methods. Recent studies emphasized ML or DL models to improve accuracy and reduce invasiveness. ML is transforming cardiovascular diagnostics in an easy way. Sevakula et al. used Random Forest (RF), Support Vector Machine (SVM) and Neural Networks to detect arrhythmia and analyze imaging [25]. Zabihollahy et al. used CNN and U-Net for detecting left ventricular fibrosis from cardiac magnetic resonance imaging (MRI) and achieved better accuracy than traditional analysis [26]. Lalli and Amutha showed that Decision Trees, k-Nearest Neighbors (KNN), SVM and Artificial Neural Networks (ANN) are used frequently for disease diagnosis and treatment prediction [27].

DL models have further improved cardiac prediction. CNNs and RNNs help classify arrhythmias and interpret ECGs [28], while reinforcement learning supports automated robotic interventions in interventional cardiology [29]. ML models can also differentiate heart failure patient outcomes effectively [30]. Ensemble methods and deep neural networks consistently outperform traditional approaches like logistic regression for CVD risk prediction [31]. NLP-enhanced ML frameworks further improve CMD diagnosis using structured and unstructured EHR data [32].

Several studies emphasized on non-invasive ways and multimodal integration. Koilpillai et al. showed that ML-assisted PET and CMR (Cardiac Magnetic Resonance) imaging also known as MRI can improve CFR estimation and it offer practical alternatives for CMD detection [33]. Georga et al. developed a multimodal ML framework which combine omics, molecular, and clinical data using random forest and multilayer perceptron (MLP) networks for CAD(Coronary Artery Disease) classification [34]. Fetit et al. showed that CNNs and gradient boosting models are applied to retinal fundus, genomic, and clinical data can predict major cardiac abnormalities in type 2 diabetes [35]. Despite these ad-

vanced and invasive methods such as CFR and the index of microcirculatory resistance (IMR) remain the gold standard for CMD detection though they are costly and invasive [17, 18]. Biomarker based studies in women with angina and no obstructive CAD showed that including inflammatory and remodeling related markers improves CMD prediction [19].

ML and DL models are increasingly used for multimodal data to improve prediction and diagnosis in CVD. Several works showed that combining features from different domains can strengthen performance noticeably. For example, multi-domain fusion of multi-channel heart sound signals has achieved high non-invasive CAD detection accuracy [36]. Ensemble models such as RF and XGBoost have also been shown to outperform traditional classifiers across common datasets [37], while ANNs demonstrated strong efficiency when compared to SVM and RF [38]. RF, in particular, performs well in situations with imbalanced datasets [39]. Research has also explored hybrid algorithms and optimization techniques to boost prediction efficiency [40], and the integration of clinical, biochemical, and metabolic data with XGBoost further improved CAD risk prediction in studies like CorLipid [41]. Feature selection and multimodal fusion continue to play an important role in building compact but accurate models [42, 43]. Broader reviews also support these findings [44] and discussed AI's role in personalized CAD care, while [45] highlighted multimodal fusion combining imaging and hemodynamic data. [46] further reviewed AI's applications in cardiovascular diagnostics, identifying interpretability and clinical validation as ongoing challenges.

More recent works have introduced advanced multimodal deep learning frameworks. Early studies focused on combining spectral heart sound characteristics with clinical metadata to support heart disease diagnosis [47, 48]. Other studies developed pseudo-siamese models to detect atrial fibrillation in noisy ECG signals [49] while frameworks using XGBoost with PCA provided interpretable detection of cardiovascular diseases [50]. Ensembles of pre-trained CNNs also achieved strong performance in ECG image classification [51] and several studies emphasized the value of combining ECG, imaging and clinical data in multimodal dataset[52]. Models like CardioNet+ fused ECG, PPG and chest X-ray data to diagnose heart failure [53] and VAE-based multimodal fusion combined X-ray and echocardiographic information for early prediction of cardiac disease [54]. Other studies used multimodal geometric learning that merges ECG with clinical text to forecast CPET outcomes [55] while combining ECG with demographic information improved myocardial infarction classification [56]. Another frameworks, such as M3S, paired Raman imaging with medical history for non-invasive CVD subtype classification [57] and ICMT-Net brought together ConvNeXt and multimodal transformers for the classification of ECG and wrist pulse signals to detect arrhythmia and coronary heart disease [58]. A multimodal dataset integrating ECG, PET/CT and stress data was also developed to support AI-based CMD diagnosis [59]. Recent reviews highlighted the grow-

ing use of ML in cardiac imaging for automated perfusion analysis and CMD detection [60].

This SLR aims to support the development of advanced diagnostic frameworks for CMD and CVD by analyzing recent research findings on the application and effectiveness of ML and DL models. This review will focus on answering the RQs set specifically for this review, as mentioned in Table 2.1 and provides insights into the potential of ML and Dl based solutions for non-invasive. accurate detection CMD.

In summary, the existing literature shows that using multimodal data along with ML and DL models is becoming essential in CVD and CMD research. Even though traditional invasive tests are still the gold standard, the newer computational and non-invasive methods are giving strong accuracy, better scalability and clearer interpretations. These approaches are helping cardiovascular care move toward more precise and patient-friendly solutions. Table 2.2 gives an overview of existing machine learning studies focused on cardiovascular and CMD detection.

Table 2.1: Research Questions for SLR and Their Rationale

ID	**RQ**	**Rationale**
SLR-RQ1	What data modalities (ECG, imaging, biomarkers, clinical records) are most frequently used in ML-based CMD prediction?	Helps to identify what kind of patient data are most commonly used in existing CMD prediction studies and how their approaches are applied in the literature.
SLR-RQ2	Which machine learning algorithms are most commonly used for CMD and CVD classification using ECG data?	Investigates which ML algorithms are mostly used for ECG-based CMD and CVD classification to understand current trends and then pick the best model for a better result.
SLR-RQ3	How does the classification performance vary between traditional ML models (LR, SVM, KNN, RF) and deep learning models (CNN, LSTM)?	Analyzes the performance differences between traditional ML and DL models to determine which group of models provides better accuracy in prediction for CMD detection.
SLR-RQ4	What role does feature selection play in improving CMD classification accuracy?	Explores how feature selection reduces dimensionality and overfitting of the ECG data, improving model generalization and highlighting the most relevant features for CMD classification.
SLR-RQ5	How effectively can ECG-only data be used as a non-invasive alternative to PET/CT for CMD detection?	Investigates whether ECG-based machine learning can serve as a reliable, low-cost, and non-invasive alternative to PET/CT imaging for CMD diagnosis, making CMD diagnosis more accessible to every stage of people.

Table 2.2: Summary of Machine Learning Studies for Cardiovascular and Coronary Microvascular Disease Diagnosis

Paper	**ML/DL Model Used**	**Dataset Used**	**Evaluation Metrics Used**	**Results**
[47]	Vision Transformer (ViT), Bio ClinicalBERT, Random Forest	PhysioNet/CinC Challenge 2022 Dataset	Accuracy, Precision, Recall, F1, AUC	Accuracy 88.1%, F1 87.7%, AUC 93.8%
[48]	PatchRWKV, TaulaTime, RF, LightGBM	Salford MINAP, SCP, Weather Dataset	Accuracy, AUC, F1, Attention	Accuracy 82.3%, AUC 0.914
[49]	KE-MPSNet (CNN, Spectrogram, Siamese)	CinC 2017, MIT-BIH AF, Chapman Dataset	Accuracy, F1, Kappa, SHAP, Grad-CAM	CinC: 82.17%, MIT-BIH: 99.11%, Chapman: 99.75%
[50]	XGBoost (Ethical ML)	UCI Heart Disease Dataset	Accuracy, Precision, Recall, F1	Accuracy 98%, Recall 100%, Precision 97%
[51]	Inception, DenseNet201, VGG16, Xception, ResNet152, EfficientNetB7	Cleveland, Hungary, Switzerland, Long Beach (ECG images)	Accuracy, F1, Precision, Recall, ROC-AUC	F1 96.51%, Balanced Acc 96.4%
[52]	CNN, ResNet, SVM, RF (Review)	–	Accuracy, AUC, F1	Comparative DL/ML analysis; highlights fusion
[53]	CardioNet+, MSTCN+Attention, ResNet-50 + ViT, SMOTE	MIMIC-III, ChestXpert	Accuracy, Precision, Recall, AUC	Accuracy 99%, strong multimodal validation
[54]	VAE, SE-Block, CBAM, Transformers	70,000+ EHR & X-rays (Columbia Univ.)	Accuracy, Precision, Recall, F1, AUC	Acc 99.52%, Recall 99.23%, DLV Acc 99.64%
[55]	Geometric Learning, TF-IDF, Riemannian Mixup	436 CHD Patients + 4000 ECGs	r-value, F1, AUC	r = 0.666 (VO_2 Peak), improved F1/AUC
[56]	xResNet-101, Multimodal Fusion	PTB-XL Dataset	Accuracy, Precision, Recall, F1, AUC	AUROC 92.1%, F1 73.5%, Spec 93.4%
[57]	Multi-scale CNN, GAF, M3S Fusion	Raman Spectra + Medical History	Accuracy, Precision, Recall, F1, Specificity	Accuracy 93.3%, F1 93.34%, Spec 97.52%
[58]	ConvNeXt, CBAM, Transformer, TL	MIT-BIH, Wuyi Pulse Dataset	Accuracy, Precision, Recall, F1	ECG Acc 81.69%, CHD Acc +6.4% (TL)
[59]	KNN, LR, RF, SVM, XGBoost	Proprietary CMD Dataset	Accuracy, Precision, Sensitivity, Specificity, F1, AUC	Acc 96.1%, Prec 100%, Sens 80%, Spec 100%, F1 88.9%, AUC 90%
[60]	Review Paper	–	–	Descriptive overview of ML-integrated imaging
[36]	SVM (RBF)	Qi Lu Hospital (36 subjects)	Accuracy, Sensitivity, Specificity	Acc 90.92%, Sens 87.96%, Spec 93.04%
[37]	KNN, MLP	UCI Heart Dataset (303 records)	Accuracy, Precision, Recall, F1, AUC	MLP: Acc 82.47%, AUC 86.41%; KNN: 73.77%
[38]	LR, SVM, NB, RF, XGB, DT, NN, kNN, GBT, RBF, MLP	UCI Heart Dataset (14 features)	Accuracy, Precision, Recall, F1	RF: 96.28%, GBT: 95.83%, MLP: 94.96%
[39]	LR, DT, RF, SVM, XGBoost, DNN	Z-Alizadeh Sani Dataset	Accuracy	DNN: 92%, XGB: 89%, SVM: 87%
[40]	LR, NB, RF	UCI Heart Dataset (Cleveland Clinic)	Accuracy, Precision, Recall, F1, ROC-AUC	RF: 86.7%, LR: 83.5%, NB: 86.7%
[45]	Review Paper	Z-Alizadeh Sani Dataset	None	Highlights ML+imaging for CAD

Continued on next page

Table 2.2 (continued): Summary of Multimodal and ML Studies

Paper	ML/DL Model Used	Dataset Used	Evaluation Metrics Used	Results
[41]	XGBoost	CorLipid (958 patients, 73 features)	AUC, Sensitivity, Specificity	AUC 0.725 (SYNTAX 23), Sens 71.1%, Spec 66.6%
[44]	Review Paper	Various Studies	None	Review—no quantitative results
[42]	Review Paper (97 studies)	Multiple Clinical Datasets	None	Overview of model validity and utility
[43]	Review Paper (3000+ studies)	EHR, ECG, CT, MRI, Omics	None	Reviews ML in cardiology; limited clinical use
[61]	DT, RF, KNN, SVM, LR, NB + FS	Cleveland Dataset (303 records)	Accuracy, Precision, Sensitivity, F-measure	DT+Wrapper: Acc 88.52%, Prec 91.3%, Sens 80.8%
[46]	RF, LR, SVM, NB	Framingham & UCI Datasets	Accuracy	RF: 84.81%, LR: 83.83%, SVM: 74.05%, NB: 54.08%
[17]	None (Review paper focused on invasive diagnostic indices)	None(summary of prior studies)	None	The paper discusses invasive diagnostic methods for CMD such as coronary flow reserve (CFR) and index of microvascular resistance (IMR). This study concludes that invasive evaluation remains the gold standard for CMD assessment but highlights the need for non-invasive diagnostic method
[26]	CNN, U-Net, GANs, RNN	Porcine & Human LGE-CMR	Dice, Precision, Accuracy, Recall	Dice 88.61%, Pixel Acc 96.98%
[27]	RF, SVM, ANN, GBM, CNN, RNN, Autoencoders	EHRs, CT, MRI, PET, Wearables	Accuracy, Sensitivity, Specificity, AUC, F1	Accuracy up to 98.8%, AUC up to 94%
[28]	CNN, RNN, GANs, K-means, AlexNet, FCN	PhysioNet, PPG, MADIT-CRT, ECGs	AUC, Precision, Recall, F1, MAE	AUC 0.93 (EF), PPV 92% (AF detection)
[29]	CNN, RL, NLP, DNN, GANs	Review Paper	Precision, Recall, Sensitivity, Accuracy	92.6% Accuracy, 96% concordance
[30]	– (Not ML-based)	431 Clinical Patients	Clinical Outcomes	10-year survival: 38%; HR: 6.0–6.8
[31]	RF, GBM, DL, XGB	UK Biobank, CPRD, Vanderbilt EHRs	C-statistic / AUROC	RF: 0.865, DL: 0.847, GBM: 0.821
[19]	LR, LASSO (Least Absolute Shrinkage and Selection Operator),RF, Extreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost)	iPOWER study dataset (1,471 women with angina and no obstructive coronary artery disease; 184 protein biomarkers measured via Olink Proteomics)	ROC-AUC (Receiver Operating Characteristic - Area Under the Curve), Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV)	Clinical model ROC-AUC: 0.61 (95% CI: 0.56–0.66),Biomarker model ROC-AUC: 0.66 (95% CI: 0.62–0.71),ML models: LASSO (0.58), RF (0.66), XGBoost (0.63), AdaBoost (0.63),, Most predictive biomarkers: Renin (REN), BNP, NT-proBNP, GDF15, ADM

Continued on next page

Table 2.2 (continued): Summary of Multimodal and ML Studies

Paper	ML/DL Model Used	Dataset Used	Evaluation Metrics Used	Results
[33]	– (Review)	Various Clinical Imaging Studies	–	Imaging-based CMD review
[34]	FFNN, SVM, RF	101 Patients (CAD and non-CAD)	Accuracy, Sensitivity, Specificity	Acc 85.1%, Sens 98.7%, Spec 44.0%
[25]	CNN, RNN, ResNet, Ensemble	PhysioNet, MIT-BIH, Mayo Clinic, UK Biobank	Accuracy, Sensitivity, Specificity, AUC, F1	AUC 0.93 (LV dysfunction), F1 0.837, Acc 91.8%
[35]	Lasso (L1 Logistic Regression)	GoDARTS (3891 patients)	Survival Curves, Log-rank p, Bootstrap	Significant stratification ($p < 0.0001$)
[18]	None (Review paper, focuses on diagnostic imaging and physiological assessment)	None(Review based on multiple studies)	None	The paper reviews diagnostic modalities like echocardiography, CMR, PET, and CT for CMD and emphasizes that despite advances, no standard diagnostic or therapeutic strategy exists. It highlights the need for multimodal approaches to improve CMD detection and treatment
[32]	SVM, RF, NB, LR, Bagging	Yale New Haven Chest Pain Center (EHRs, PET/CT)	Precision, Recall, F1	F1: 0.81 (CMD vs CAD), 0.74 (CMD vs non-CAD)

2.1.1 Method

The study explores how CMD or other heart diseases can be detect using machine learning and deep learning models. We conducted a systematic review on previous research papers according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). In the clinical sector, the use of PRISMA have become increasingly important [62]. The PRISMA guidelines are shown in Figure 2.1. A systematic search was done using PubMed, IEEE Xplore, and the ACM Digital Library. The search was limited to studies published from 2019 to 2025 in English. These databases provided research papers from multiple authors which were screened for relevance based on our eligibility criteria.

2.1.2 Search Strategy

First of all, we identified the keywords that are relevant to our topic.The search string used was: (“Coronary Microvascular Disease” OR “Heart Disease”) AND (“Machine Learning” OR “Deep Learning”). Then we entered the resulting search string that consists of keyword combinations using the Boolean operators “AND” (must be included) and “OR” (may or may not be included). Initially we restricted our search to the papers’ titles, abstracts and the keywords. The initial search gave us 3670 papers, from which

we have only included the ones that strictly aligns with out study. Table 2.3 shows an overview of the inclusion and exclusion criteria applied to select the best studies for the systematic literature review.

Table 2.3: Lists of Inclusion and Exclusion Criteria

Area Specification	**Inclusion Criteria**	**Exclusion Criteria**
Database	PubMed, IEEE Xplore, and the ACM Digital Library.	Databases not included in the selected search sources.
Journal/Conference Type	Research article, Research review article	Book chapters, short papers, abstracts etc.
Area of Interest	Coronary Microvascular Disease, Heart Disease, Machine Learning, Deep Learning,	Areas excluding "Inclusion Criteria"
Selected Time Period	2019-2025	Before 2019

Inclusion Criteria

We specifically focused on the diagnosis, prediction, or assessment of CMD or other microvascular cardiac disorders using ML or DL models. We got 2948 articles after excluding non-relevant abstracts. We read the title and abstract manually to assess the relevance to the inclusion criteria. It reduced into 40 articles those were considered for inclusion in this systematic review. These studies were added only if they followed the criteria mentioned below:

- Papers published between 2019-2025
- Focused on heart disease specially CMD and CVD
- Studies must diagnosis, prediction, or assessment of CMD using ML or DL models

Exclusion Criteria

The exclusion criteria that were followed to exclude a study are:

- Papers published before 2019
- Duplicate studies or papers that reported overlapping results
- Studies are not related to the topic

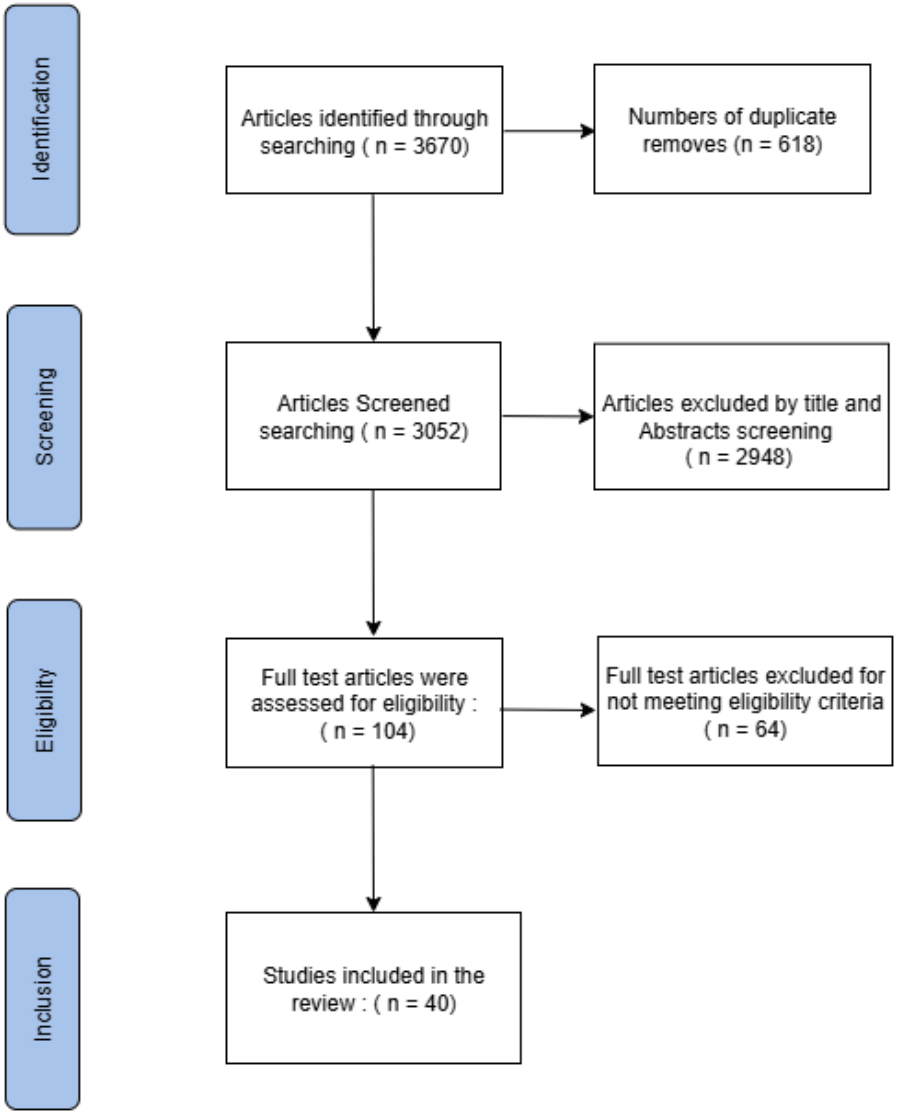


Figure 2.1: PRISMA flow diagram of the literature search and selection process.

Study Selection

A total of 3670 articles were initially retrieved from major databases, including PubMed, IEEE Xplore, and the ACM Digital Library. After importing the search results, 618 duplicate records were identified and removed using an automation tool. Following this step, 3052 articles remained for title and abstract screening. Out of these, 2948 articles were excluded for being unrelated to CMD, lacking machine learning or deep learning components, or not meeting the predefined inclusion criteria, lack the depth and detailed analysis required for this study.
This left 104 articles for full-text assessment. After reviewing the complete manuscripts, 64 studies were excluded because they did not provide sufficient model details, focused on objectives outside the scope of this systematic review and they did not meet the inclusion criteria outlined in Table 2.3. In the end, 40 studies met all inclusion criteria and were selected for the final analysis. The entire study selection process adhered to the PRISMA guidelines and is visually summarized in Figure 2.1.

2.1.3 Data Extraction

This phase is focused on the in-depth analysis of the selected 40 articles to identify the exact relevant data based on the systematic review. As shown in Figure 2.2, the data extraction themes help to identify the main research themes addressed in the reviewed studies.

Topical Association

- Title Based: Initially, the topical screening was performed by analyzing the paper titles. If the title was relevant to the diagnosis or detection of CMD or CVD or heart disease using machine learning or deep learning.
- Keyword-based: Analyzed the keywords given by the authors used in each articles. The keyword were used to generate a word cloud showed in Figure 2.3 which highlights the major keywords and displays the frequently used terms in the selected research papers.

Publication Profile

- Publication Type: All the articles were classified by the publication type including journal articles, conference articles.
- Publication Year: To observe the condition and growth of research in the domain, we selected articles from 2019 to 2025.

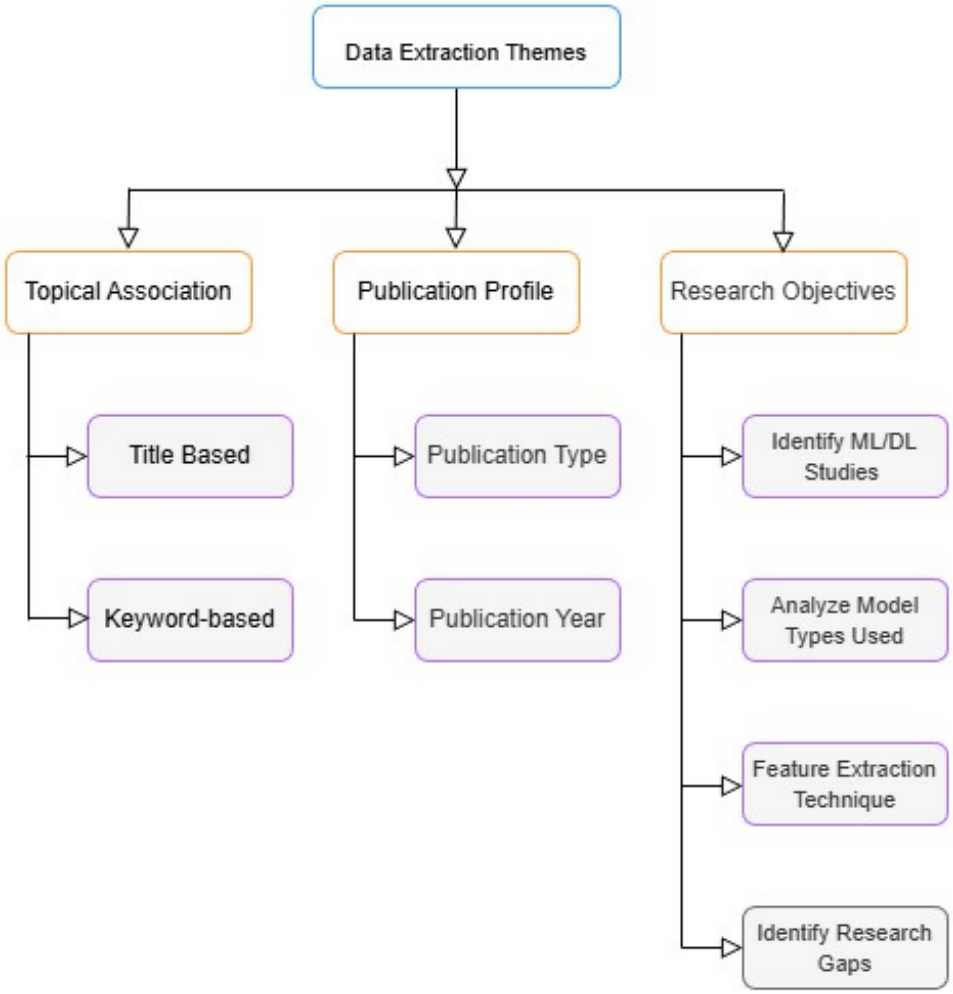


Figure 2.2: Data Extraction Themes

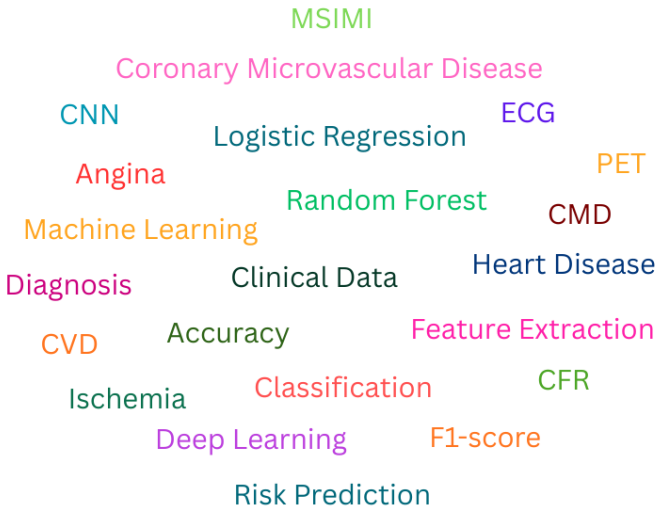


Figure 2.3: Word cloud based on the keywords.

Research Objectives

The research objective of this SLR is to know which the machine learning and deep learning models are used and how these models have been applied to detect, diagnosis and predict CMD and for other heart diseases.

- Systematically identify and compile peer-reviewed studies (2019–2025) that utilize ML or DL techniques for CMD or other Cardiovascular Disease (CVD) diagnostics.
- Analyze the model types, feature extraction techniques, multimodal data integration, and evaluation strategies used in these studies.
- Identify the research gaps and future works.

2.1.4 Literature Search Outcomes

Following the PRISMA guidelines, a detail literature research was conducted across three major scientific databases: PubMed, IEEE Xplore, and the ACM Digital Library. After the initial search we found a total of 3,670 records. After removing 618 duplicate entries in several databases, 3,052 unique articles remained for title and abstract screening. During the abstract screening phase, 2,948 studies were excluded because they did not meet the pre-defined inclusion criteria. The major reasons for exclusion were irrelevance to CMD, absence of machine learning or deep learning use, and less technical depth. After the exclusion, 104 articles were selected for full text assessment. After a detailed full text

review, 64 studies were further excluded due to inadequate methodological descriptions or lack of ML or DL-based diagnostic focus. Finally, 40 studies published between 2019 and 2025 met all the inclusion criteria and were selected for the final systematic analysis. These selected studies form the main knowledge base for understanding current research trend on using machine learning and deep learning models for diagnosing CMD and heart disease. The results show that researchers are focusing on combining different types of data, using safe and non-invasive diagnostic methods more, and applying advanced deep learning models. But the limitations were also stated in each study to address them in any future researches. Table 2.4 summarizes the common limitations found while doing the literature search.

Table 2.4: Summary of Common Limitations Reported in CMD ML/DL Studies

Limitation Category	**Description**	**Examples in Studies**
Limited Sample Size	Small datasets cause overfitting and poor generalization	Single-center studies, low CMD cases
Lack of Multimodal Data	Use of only one data type	Only ECG, only imaging, no biomarkers
Imbalanced Datasets	Uneven CMD vs healthy samples	CMD under-represented
Limited Clinical Validation	No real-world or multi-center testing	Only internal validation
Data Quality & Noise Issues	Poor signal quality or missing data	Noisy ECG, inconsistent images
Limited Feature Engineering	Weak preprocessing or feature selection	No feature selection techniques applied
Overfitting in Complex Models	Deep models trained on small datasets	CNN overfitting
Lack of External Benchmarking	No comparison with baseline models	No logistic regression baseline
Population Bias	Limited demographic diversity	Only European or East Asian subjects
High Computational Cost	Large models require heavy computation	GPU-dependent CNN training
Limited Public CMD Data	No open CMD datasets available	Private hospital datasets

2.1.5 Findings

SLR-RQ1: What data modalities (ECG, imaging, biomarkers, clinical records) are most frequently used in ML-based CMD prediction?

- In the article number [33, 49] indicates that multimodal data has become central to CVD and CMD prediction. ECG signals and clinical data are widely used methods in CMD and CDV studies.
- According to article [19] Biomarkers especially in women with angina significantly enhanced CMD prediction and were included in several multimodal studies.S
- From article [35] can describe that imaging sources such as retinal fundus photographs can contribute to prediction of cardiac abnormalities.

SLR-RQ2: Which machine learning algorithms are most commonly used for CMD and CVD classification using ECG data?

- Traditional ML models like SVM, RF, Decision Trees, KNN, LR, XGBoost are widely used for ECG derived feature classification [25, 27].
- Deep learning architectures especially CNNs, RNN and LSTM are the most common models for raw ECG waveforms [28].
- Some recent studies used transformers and multimodal deep learning combined with ConvNeXt to extend ECG analysis by integrating pulse signal to improve the accuracy of diagnostic [58].

SLR-RQ3: How does the classification performance vary between traditional ML models (LR, SVM, KNN, RF) and deep learning models (CNN, LSTM)?

- When dataset is imbalanced machine learning models specially RF perform better [37].
- Combination of deep learning and machine learning models often give better results, can improve prediction accuracy over unimodal models [53, 54].
- CNNs and RNNs or LSTMs performs better than traditional ML models in arrhythmia detection and ECG interpretation. These models can automatically extract complex features from the data [28, 31].

SLR-RQ4: What role does feature selection play in improving CMD classification accuracy?

- Feature selection reduces overfitting, especially important for small CMD datasets. Feature selection is important to select the important features to develop models and clinical, biochemical, and metabolic features can significantly enhance CAD prediction performance shown in [41, 42].
- In the article [47, 48] explains that feature extraction from heart sound signals become strengthen to non-invasive CAD detection accuracy.

SLR-RQ5: How effectively can ECG-only data be used as a non-invasive alternative to PET/CT for CMD detection?

- In article [17, 33] PET/CT and MRI tests measures CFR amd IMR. These are the more accurate methods for CMD diagnosis. But ECG can only measure the heart signals. So ECG is helpful but cannot replace PET/CT.
- ECG cannot fully replace PET/CT but it can offer a non-invasive, low-cost alternative for CMD detection [53, 55, 59].

Chapter 3

Methodology

3.1 Dataset Description

The dataset used in this study was developed by a multidisciplinary research team and collected from the Department of Cardiovascular Imaging and Nuclear Medicine at China [21]. This is a publicly available dataset which is designed to facilitate biomarker discovery and machine learning based screening for CMD. It primarily focuses on ECG-based features received during pharmacological and mental stress tests to improve CMD detection, also to detect CMD from similar cardiovascular disorders.

A total of 120 female participants aged between 18 and 75 years from the Guangdong Provincial People's Hospital, China, shared their health data. Among them, 80 participants were angina patients without obstructive coronary artery disease (ANOCA) and 40 age-matched healthy controls. Excluding signal issues or adenosine intolerance, the dataset includes 97 participants. Figure 3.1 gives an overall overview of the dataset.

All participants underwent adenosine-induced stress testing over 6 minutes using a standardized infusion protocol of 140 kg/min. ECG recordings were collected in 3 stages. They were rest, stress, and recovery. Besides, to assess myocardial blood flow and determine coronary flow reserve (CFR), PET/CT imaging using a 13N-ammonia tracer was used during the stress stage [59].

The dataset files are organized by participant ID, with ECG data stored in .csv format and corresponding PET/CT and diagnostic labels in JSON format. Each participant's directory includes ECG data for all three stages, along with the associated diagnostic labels for CMD and MSIMI. The dataset also includes derived features, data processing scripts, and metadata files describing the experimental protocols and diagnostic criteria [21].

A total of 88 ECG-based variables were extracted from each stage, as well as heart rate variability indices. Similarly, each participant shares 264 ECG variables corresponding to the three physiological stages. These features were used to train multiple machine

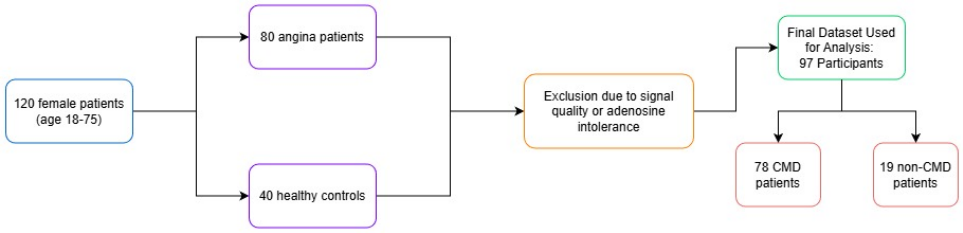


Figure 3.1: Overview of the dataset.

learning models: LR, KNN, RF, SVM, and XGBoost, for CMD classification.

3.2 Data Pre-processing

The dataset consists of 97 female participants, excluding data with signal issues or adenosine intolerance. Several patients were excluded due to missing or incomplete data. The final dataset with 97 subjects has the complete physiological, biochemical, and psychological data for each individual. All participants went through an ECG during three physiological states: Rest, Stress, and Recovery, each a 6-minute phase, along with PET/CT imaging for CFR [21]. CFR is the increase in blood flow in the heart when it is above its resting rate. In other words, CFR shows the ability of the coronary arteries to respond to the heart's oxygen demand during heavy activities or exercises [63].

For all subjects in the dataset, their CMD is labeled as 1 (CMD positive) or 0 (CMD negative), based on the CFR value, where:

- CFR < 2.0, indicates positive CMD
- CFR ≥ 2.0, indicates negative CMD

For the ECG record, the dataset includes 18 minutes of 12-lead signal collected at 500 Hz with 16-bit precision for each subject. The signal was divided into three equal phases of 6 minutes each. For the stress phase, all participants underwent an adenosine stress test [21]. Adenosine stress is a commonly used process to increase the heart rate. It is a medical method to increase the stress on the heart, especially used for patients who cannot perform physical exercises [64]. Because of this, the stress phase shows more muscle activation noise. Thus, this stage in particular contains more noise than the other stages.

Other than this, ECG is also highly sensitive to distortions, which can lead to inaccurate analysis. Therefore, to make the ECG signals suitable for machine learning, three primary noise sources were identified from which most of the noise might be coming.

- **Powerline Interference:** An electrical noise of around 50–60 Hz caused by electrical power lines.

- **Baseline Wander:** A low-frequency disturbance, not of cardiac origin, for example, breathing or physical movement.
- **Muscle Contraction:** This noise is caused by unconscious muscle movement, for example, shivering from anxiety or stress.

To remove these noises, a two-stage median filtering process was applied in combination with the Daubechies wavelet.

- **200 ms Median Filter:** A pre-processing step used to remove sharp and fast noise.
- **600 ms Median Filter:** This filter removes baseline wander or low-frequency noise. It smooths out the waves, creating a clean signal that makes it easier to perform further analysis.

Following the median filtering, the Daubechies wavelet level 6 decomposition was applied for more precise noise removal. The Daubechies wavelet level 6 is widely used in ECG processing as the shape is very close to the ECG waves. It is decomposed into several levels where it removes high-frequency noise, corrects low-frequency baseline wander, and finally keeps the clean ECG morphology.

Now that the noise is removed and the clean waves are detected, both morphological features and HRV indices were extracted using NeuroKit2. Before extracting the features, all the signals were cleaned using its built-in methods. NeuroKit2 is an open source, user-friendly package used for neurophysiological signal processing that helps to process different body signals like ECG, EDA, EMG, and breathing [65]. It uses built-in pipelines that can detect key features.

All signals were preprocessed and cleaned using the ecg_clean function from the NeuroKit2 library to remove noise. After cleaning the data, feature extraction was performed using NeuroKit2's built-in ECG feature extraction functions. During feature extraction, ECG features could not be successfully extracted for three subjects (two from CMD positive and one from CMD negative). Therefore, these three subjects were excluded. Now, the final dataset consists of 94 subjects (77 non-CMD and 17 CMD). For every ECG lead, 105 physiological features were extracted. These included morphological features and HRV features generated from the ECG waveforms. NeuroKit2 generated 3 features that contained NaN values consistently during extraction for all the subjects. These features were removed to ensure data quality and model stability. After removing the 3 NaN features, the final feature count became 102 features. This resulted in an initial feature matrix of size (94 subjects $\times$ 12 leads) $\times$ 102 features, meaning 1128 rows $\times$ 102 columns. These 12 rows per subject cannot be treated as independent observations because they belong to the same person. Therefore, a feature-level aggregation strategy was applied

to combine the 12 leads into representative values for each subject. For each subject, the 12-lead signal feature set was transformed into a single signal by computing:

- Mean value of 12 leads for each feature.
- Standard deviation of 12 leads for each feature.

This produced 2 aggregated values for each feature. Therefore, the initial 102 features expanded to 102 mean features and 102 standard deviation features, 204 aggregated features per subject in total. After cleaning and aggregation, the feature matrix used for classification had dimensions of 94 subjects $\times$ 204 features. In addition to the aggregated features, we obtained 204 features separately for the rest, stress, and recovery stages for each subject. This resulted in three sets of 204 features for each subject. To capture the physiological changes between stages, two delta scores (stress $-$ rest and recovery $-$ stress) were calculated. After computing the features of rest and the delta scores, the total number of features became 612 for each of the 94 subjects. Now, the final dataset dimension is 94 subjects $\times$ 612 features. This dataset represents the combined statistical behavior of all 12 ECG leads for each individual, making it suitable for machine learning classification of CMD.

3.3 Feature Selection

After creating the final dataset, having 94 subjects and 612 ECG-derived features, feature selection is the next step before developing any machine learning model. Feature selection is an important part when building models, because in biomedical signal analysis, datasets mostly contain a large number of variables, and it is very rare that all the variables are useful. The purpose of feature selection is to find the relevant, informative, and non-redundant features from the entire feature set. Removing noisy, redundant, and weakly informative features improves models' accuracy, reduces overfitting. In this study, the goal is to identify ECG-derived features that most effectively differentiate CMD-positive subjects from CMD-negative ones. For feature selection, two approaches were used to select features for CMD detection.

- **Delta Score:** In addition to combining 12 ECG leads, the study also sought to understand how the heart's behavior changed during three phases: rest, stress, and recovery. For this, the same 204 aggregated ECG features were extracted separately for each stage. Two delta scores were computed for each subject. The first one was stress, the second one was recovery.

    1. **Stress Delta Score:** Stress score was calculated by subtracting the Rest values from the Stress values. This shows each feature changes when the heart goes from the rest state to the high stress state caused by adenosine.

2. **Recovery Delta Score:** This value was calculated by subtracting the Stress values from the Recovery values. This explains how the features change when the heart returns to normal after the high stress state.

The final feature matrix was created by combining these 204 Rest features with the two delta-score sets, which gave a total of 612 features per subject. The delta score is important because it shows how the heart reacts and adapts to stress, and it can observe the change in the heart's condition.

- **Mutual Information (MI):** Mutual information is a measure from information theory that describes how much information one variable provides about another one. It checks how strongly a feature and the output label are connected. For feature selection it helps to know which features in a dataset are most important for predicting the target label. Mutual information can detect both linear and nonlinear relationships. In this study, the final 94×612 feature matrix is enough, making feature selection a necessary step to reduce dimensionality. For each of the 612 features, mutual information was calculated. Features with higher MI scores are considered more relevant for the classification of CMD. Later on, this process selects a smaller subset of the informative features.

3.4 Model Development

3.4.1 Logistic Regression

Logistic Regression (LR) is a multivariable statistical analysis method used to model the relationship between multiple independent variables and a single categorical dependent variable by fitting data to a logistic curve [66]. There are two types of LR: binary logistic regression and multinomial logistic regression. Binary logistic regression is typically applied when the dependent variable is dichotomous, and the independent variables are either continuous or categorical. The model estimates the probability (p) of an event's occurrence by fitting data to an 'S'-shaped logistic curve. The parameters of the model are estimated using the Maximum Likelihood Estimation (MLE) method. LR works by transforming the probability into log odds, which is then modeled as a linear function of the predictor variables. The key features and clinical utility of LR are described below:

- **Independent Variables:** Age, sex, cholesterol level, resting blood pressure, chest pain type, fasting blood sugar, and maximum heart rate achieved.
- **Interpretation via Odds Ratios (OR):** The OR quantifies the effect of each predictor on the likelihood of Coronary Heart Disease (CHD). OR > 1 indicates increased risk, while OR < 1 indicates a protective effect.

- **Risk Stratification:** LR allows estimation of individualized CHD probabilities, supporting early screening, prioritization of diagnostic tests, and preventive interventions.

The transparency and clinical interpretability of the LR model make it particularly suitable for the prediction of coronary heart disease [66], providing actionable information on the contribution of multiple risk factors to disease development.

3.4.2 Support Vector Machine (Kernel - RBF)

Support Vector Machine (SVM) is a supervised machine learning algorithm which is used for classification and regression tasks. The SVMs provide a compromise between the parametric and the pure nonparametric approaches: As in linear classifiers, SVMs estimate a linear decision function, with the particularity that a previous mapping of the data into a higher-dimensional feature space may be needed [67]. Its main objective is to identify the best boundary, known as a hyperplane, that separates different classes in the data. It is effective for binary classification, such as spam vs. not spam or cat vs. dog. The main goal of SVM is to maximize the margin between the two classes because a larger margin gives better performance on new and observed data.

- **Hyperplane:** A Hyperplane is a boundary which separates different classes in feature space, and it is represented by the equation wx + b = 0 in linear classification, where w is the weight vector, and b is the bias term.
- **Support Vectors:** The closest data points of the hyperplane are called support vectors. These points are crucial for determining the hyperplane, and the margin is measured from these points in SVM.
- **Margin:** The margin is the distance between the hyperplane and the support vectors. The main aim of SVM is to maximize this margin for better classification performance.
- **Hard Margin:** Hard Margin is a maximum-margin hyperplane that perfectly separates the data without misclassifications.
- **Soft Margin:** Allows some misclassifications by introducing slack variables, balancing margin maximization and misclassification penalties when data is not perfectly separable.
- **Kernel:** A kernel function maps data to a higher-dimensional space to handle data which are not linearly separable. This transformation allows SVM to create a linear separation in the transformed space. By applying a kernel function, SVM transforms the data points into a higher-dimensional space where they become

linearly separable. Radial Basis Function (RBF) Kernel transforms data into a space based on distances between data points. The RBF kernel, also known as the Gaussian kernel. The RBF kernel implicitly maps input data into an infinite-dimensional feature space, allowing the model to capture highly complex, non-linear relationships among features. The final decision function in this feature space is expressed in terms of the kernel and Lagrange multipliers obtained from the dual optimization problem [67].

3.4.3 Random Forest

Random Forest (RF), a machine learning model, uses multiple decision trees to make the predictions better. Each decision tree looks at different random parts of the data to produce a result. Each tree gives its own prediction based on the part of the data. For classification, all the results are combined by voting, and for regression, the results are combined by averaging, which is why it is known as an ensemble learning technique. This process helps avoid overfitting, thus making the model's prediction more accurate. In classifying CMD, Random Forest is a critical algorithm as it can handle missing data, shows feature importance to make useful predictions, and also works well with complex data [68].

3.4.4 K - Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) is a supervised classification algorithm which works by finding the closest data points. It finds "k" closest data points or neighbors to the given input to make predictions. "k" here is the number that tells the algorithm how many neighbors or nearby points to look at for making the decision. KNN operates on the similarity of the nearest data points and predicts the new data point by considering the nearest ones in the training dataset. The value of "k" should be picked based on the data to get the best possible outcome. Several statistical methods are used to select the value of "k" [69].

- **Cross-Validation:** The K-fold cross-validation is one of the good ways to find the best value of "k". K-fold means dividing the dataset into "k" parts and then training some of them and testing on the others. This is a repeated process used on each part, which means every subset of the dataset is used both as a training part and as a testing part. The "k" value with the highest average accuracy is then selected as the best one to use.
- **Elbow Method:** A graph is drawn to show the error rate or accuracy for different "k" values. The value of "k" and the error rate are inversely proportional in the beginning. When "k" increases at first, the error rate decreases till a certain point,

and after that it stops decreasing quickly. That exact point where it looks like an "elbow" is usually chosen as the best value of "k".

- **Odd Values for K:** Especially in classification problems, it is a common idea to use an odd number for "k", so that there is no hesitation using the most common class among the neighbors, and the middle one can be easily chosen.

For classification, KNN makes predictions based on the majority class, and for regression, it makes predictions based on the average value. KNN is also known as a non-parametric and instance-based learning model, as it only predicts using nearest data points and not the underlying data distribution. There are multiple metrics used in the KNN algorithm [69].

- **Euclidean Distance:** This is a straight-line distance or the shortest path between two points in a plane or space.
- **Manhattan Distance:** Manhattan or Taxicab distance is the total distance calculated using only the horizontal and vertical lines, like a grid.
- **Minkowski Distance:** It is a generalized metric that unifies several well-known distance metrics, such as Euclidean or Manhattan distance, depending on the value of P.

3.4.5 XGBoost

XGBoost is an ensemble learning method that can achieve high accuracy on complex datasets. It is an optimized distributed gradient boosting. This algorithm uses a decision tree as a base model and combines them. Each new tree corrects the mistakes of previous ones. This process is known as boosting. Boosting helps to improve the performance. As it runs tasks simultaneously, it can work fast and train models on large datasets. At first, the model is trained on the data and averages the target variables. Then it calculates the error between the predicted and actual values. After that, it trains the next trees based on the errors of the previous tree. This process continues until the model stops improving. The final prediction is the sum of all predictions [70].

3.4.6 MLP - MultiLayer Perceptron

A multi-layer perceptron (MLP) is a type of feed-forward Neural Network with multiple neurons arranged in layers [71]. This includes an input layer, hidden layers, and an output layer. Each layer contains a set of perceptron elements, which are called neurons. All the neurons in a layer are fully connected to the neurons in the next layer. There is no cycle or loop.

- **Input Layer:** In this layer, each neuron is associated with an input feature.
- **Hidden Layer:** MLP can contain any number of hidden layers that can contain any number of neurons or nodes. This layer receives information from the input layer and processes it.
- **Output Layer:** This layer processes the result. If there are multiple outputs, the layer will have the same number of neurons.

The key mechanisms are:

- **Froward Propagation:** The data transfer from the input layer to the hidden layer to the output layer. In the hidden layer, each neuron processes data by computing a weighted sum, and this weighted sum is passed through an activation function. This activation function introduces the non-linearity.
- **Loss Function:** After generating an output, it needs to calculate the loss function.
- **Backpropagation:** To minimize the loss function, gradient calculation, error propagation, and gradient descent are used.
- **Optimization:** MLP depends on optimization algorithms using some methods to clarify the weights and biases during training.

3.4.7 Ensemble Learning

The ensemble model was developed by combining three powerful models: SVM, Random Forest, and XGBoost. Each model has different strengths for prediction. We utilize the "Soft Voting" method to develop these models. In this strategy, each model does not just need to give a "Yes" or "No" answer, but needs to provide a probability score that shows how it can predict confidently. This strategy takes these probability scores and calculates the average. This score is used for the final prediction. By combining all the outputs from three models, it reduces the impact of any mistakes. To make more reliable predictions, an ensemble model works better. Because if any model makes any mistake, another model can balance it out through their probability scores. It helps to predict more stable and accurate CMD detection than any single model.

3.5 Evaluation Metrics

The performance of the CMD prediction models was assessed using accuracy, precision, recall, F1-score, specificity, and the confusion matrix. These metrics are essential for understanding how reliably the machine learning models can detect the presence of

CMD. In the equations below, TP, TN, FP, and FN refer to True Positive, True Negative, False Positive, and False Negative predictions, respectively.

- **Accuracy:** Accuracy represents the ratio of correctly predicted instances to the total number of instances. A perfect model can have zero false positives and zero false negatives. In the case of CMD classification, high accuracy means the model is correctly identifying both CMD and non-CMD cases. The equation for accuracy is as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad [72]$$

- **Precision:** Precision is the ratio between the True Positives and all the positives. It helps to understand how accurate a model's "yes" predictions are. It measures the proportion of correctly identified CMD cases among all instances predicted as CMD. High precision is essential in CMD because false positives could lead to unnecessary stress testing, interventions, or misdiagnosis of patients without CMD.

$$\text{Precision} = \frac{TP}{TP + FP} \quad [72]$$

- **Recall:** Recall can also be called True Positive Rate (TPR) or Sensitivity. It represents the percentage of actual positive cases that the model correctly predicts. A high recall value means the model is able to detect most CMD cases. The equation for recall is as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad [72]$$

- **F1-Score:** The F1-score is the harmonic mean of precision and recall, which combines both metrics into a single value. It balances the importance between false positives and false negatives, which is particularly important in CMD datasets where the number of actual CMD patients is smaller than controls. A high F1-score indicates that the model effectively identifies CMD patients without missing too many cases or over-predicting.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad [72]$$

- **Confusion matrix:** The Confusion Matrix is used to measure how well a classification model performs. It compares the predictions made by the model with the actual results, showing the counts of TP, TN, FP, and FN. For CMD, it allows us to see exactly how many patients were correctly identified, how many CMD cases

were missed, and how many controls were incorrectly classified as CMD. [73]

Actual / Predicted	**Predicted CMD**	**Predicted non-CMD**
Actual CMD	TP	FN
Actual non-CMD	FP	TN

These metrics collectively offer a detailed understanding of model performance, highlighting both its strengths in detecting CMD and its limitations in avoiding misclassification, which is critical for safe and accurate clinical application.

3.5.1 Cross Validation

Cross-validation (CV) is a method used to check how well a machine learning model performs on unseen data. In this study, we have used Leave One Subject Out CV. In this method, the model is trained on the whole dataset except for one data point. This point is used for testing, and this process is continued for every data point, resulting in low bias. It can cause high variance, and it can also be very time-consuming for large datasets. This method is also computationally intensive, especially for large datasets. It requires as many iterations as there are data points.

Chapter 4

Results & Analysis

4.1 Logistic Regression

The Logistic Regression model was evaluated using the confusion matrix and different classification metrics to understand how effectively it can distinguish between CMD and non-CMD patients using ECG features. The confusion matrix Figure 4.1 shows that the model correctly identified 51 non-CMD subjects, and 26 non-CMD subjects were incorrectly classified as CMD. This shows that the model performs better when detecting non-CMD or healthy patients. But for actual CMD patients, the results were weaker. Only 9 CMD cases were correctly predicted as CMD, while 8 CMD patients were mistakenly classified as non-CMD. This indicates that the model struggles to identify CMD positive patients.

From Table 4.1, the overall performance of the LR model can be clearly observed. The overall accuracy of the LR model was 64%, meaning that the model correctly classified about 64 out of 100 patients. The macro precision was 0.56, which represents the average precision value of CMD and non-CMD classes. It shows the moderate overall reliability of the model's predictions. However, the macro recall was only 60%, which is very low for medical diagnosis. This means the model was able to detect only around half of the actual CMD patients, missing the remaining cases. Missing CMD cases is a serious issue because CMD is a cardiovascular condition that requires timely attention. The model is better at identifying healthy non-CMD patients but performs poorly in finding true CMD patients. The macro F1-score of 0.55 reflects the imbalance between precision and recall and further indicates that the model is not stable enough for reliable medical decision-making.

Overall, the LR model shows better performance for identifying non-CMD but struggles with detecting CMD patients.

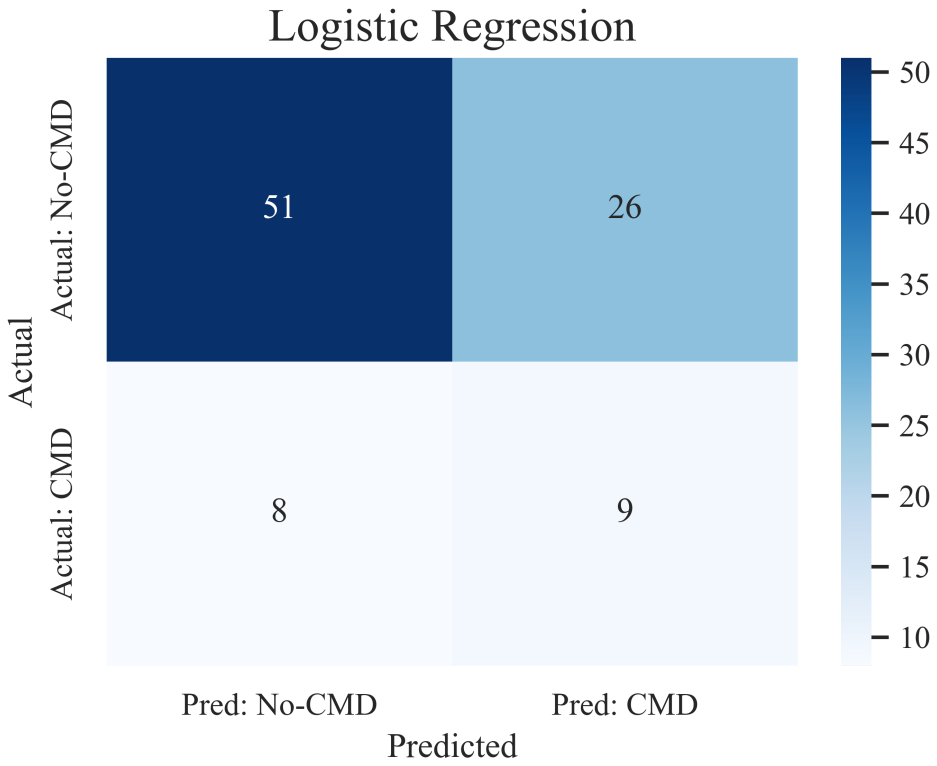


Figure 4.1: Confusion Matrix for Logistic Regression

4.2 Support Vector Machine (Kernel - RBF)

The SVM model using the RBF kernel showed a comparatively weaker performance than the other models in this study in distinguishing between CMD and non-CMD patients. The confusion matrix shown in Figure 4.2 demonstrates that the classifier correctly predicted 66 non-CMD patients and 7 CMD patients. However, 11 non-CMD cases were misclassified as CMD and 10 CMD cases were incorrectly labeled as non-CMD. This misclassification pattern indicates that the model struggles to consistently differentiate between the two classes particularly for CMD positive cases.

From Table 4.1, the overall performance of the SVMs (Kernel - RBF) model can be clearly observed. The model achieved an overall accuracy of 78%, which demonstrates a moderate level of agreement between predicted and true labels across the dataset. The macro precision score of 0.63 shows that the model is correct in about two-thirds of its CMD predictions. Meanwhile, the macro recall of 63% indicates that the model captures roughly two-thirds of true CMD cases, which is relevant for clinical screening tasks. Finally, although the macro F1-score of 0.63 shows a moderate balance between precision and recall, the SVM (RBF Kernel) model still performs relatively poorly compared to other models in this study.

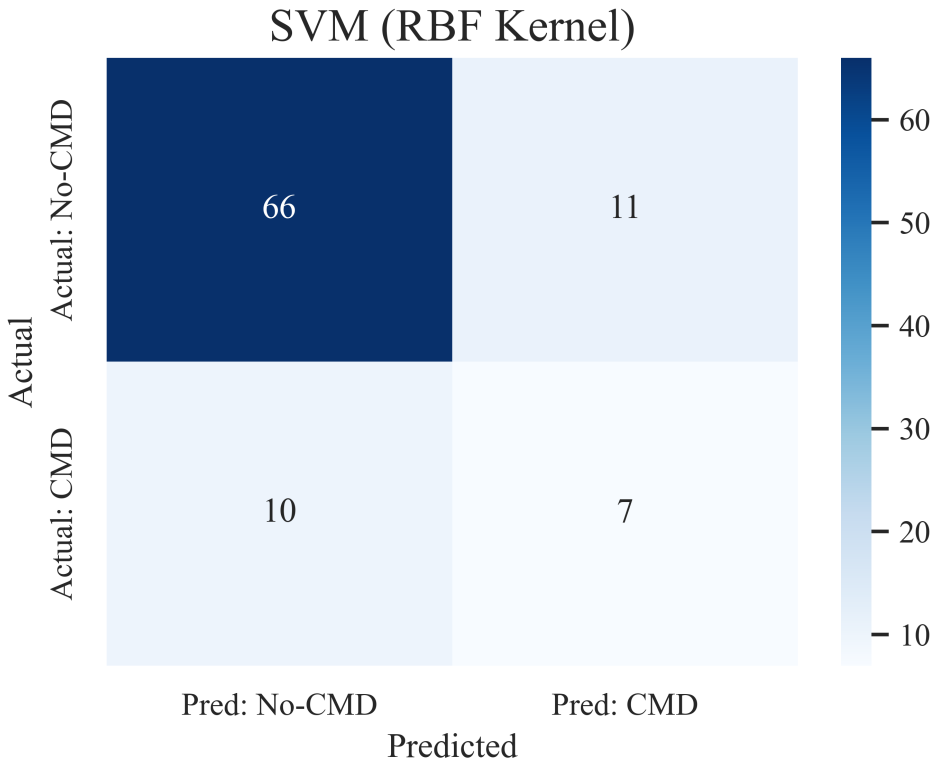


Figure 4.2: Confusion Matrix for SVM (RBF Kernel)

4.3 K-Nearest Neighbor (KNN)

The performance of the KNN model was evaluated using the confusion matrix and standard classification metrics to understand how well ECG-derived features can distinguish between CMD and non-CMD patients. For our study we used the value 'k' = 5, which means considering 5 nearest data points to predict the result.

From the confusion matrix in Figure 4.3, it was observed that 45 non-CMD patients were correctly classified as non-CMD, while 32 non-CMD patients were misclassified as CMD. On the other hand, out of all actual CMD patients, 15 were correctly detected as CMD, while 2 CMD patients were wrongly classified as non-CMD. This clearly shows that the KNN model performs much better in identifying CMD patients compared to detecting non-CMD patients.

From Table 4.1, the overall accuracy of the KNN model was 64%, which indicates that the model correctly classified around 64 out of 100 patients. The macro precision of the model was 0.64, meaning that when the model predicted a patient to be CMD positive, it was correct about 64% of the time. This indicates a moderate level of reliability for positive predictions.

However, the macro recall was 73%, which means that the model was able to correctly identify about 73% of the patients, while the remaining subjects were missed.

The macro F1-score of 0.60 further shows the imbalance between precision and recall,

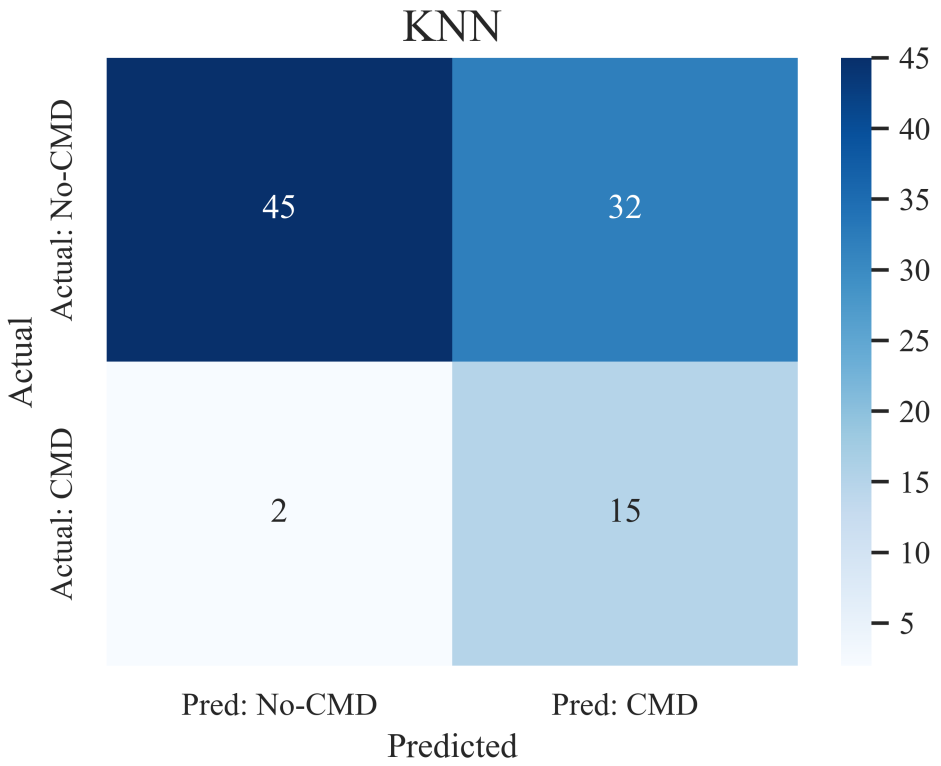


Figure 4.3: Confusion Matrix for KNN

showing that the model lacks stability in detecting CMD reliably.

The poor result of KNN is likely due to class imbalance, overlapping ECG feature patterns between CMD and non-CMD patients, and the sensitivity of KNN to local feature distributions. KNN can be useful as a baseline model, but it is not reliable enough for independent CMD classification using ECG data as it highly misclassify the non-CMD patients.

4.4 Random Forest

The RF model showed significantly stronger performance compared to a lot of other models in detecting CMD using ECG-derived features.

From the confusion matrix in Figure 4.4, 71 non-CMD patients were correctly classified as non-CMD, and 6 non-CMD patients were incorrectly classified as CMD. More importantly, 7 CMD patients were correctly detected, while 10 CMD patients were misclassified as non-CMD.

From Table 4.1, the accuracy of the RF model reached 83%, meaning that 83 out of 100 patients were classified correctly. This accuracy indicates good overall performance. The macro precision was 0.71, which means that when the model predicted a class, it was correct 71% of the time. This level of precision is impressive in the critical medical sector. The macro recall of the RF model was 67%, which is extremely important for

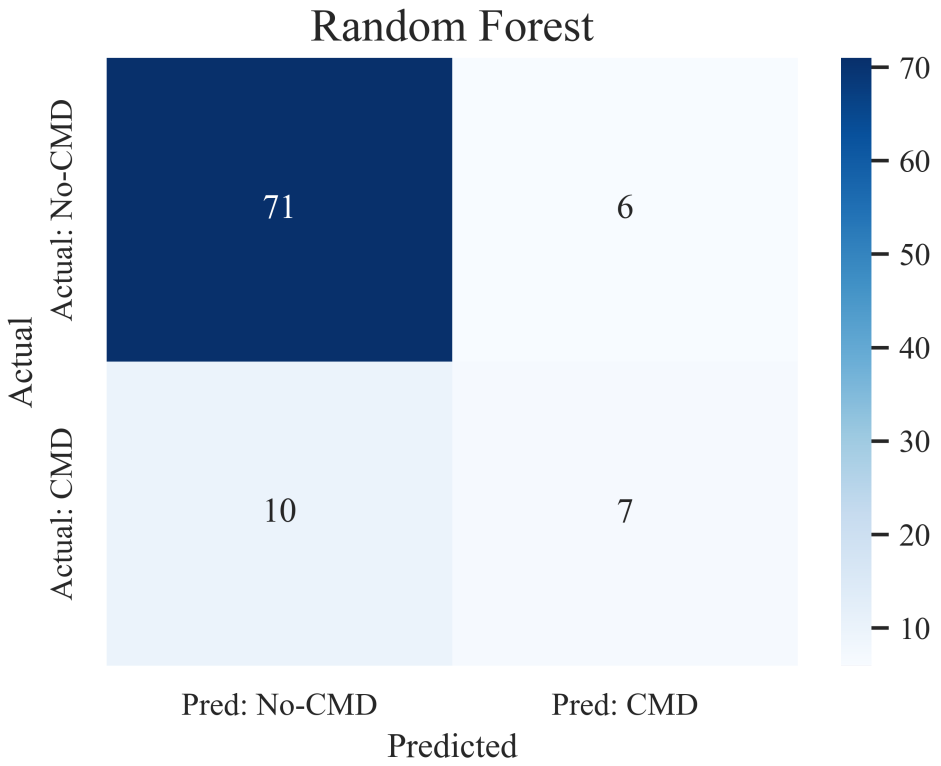


Figure 4.4: Confusion Matrix for Random Forest

CMD detection. This means that 67% of the patients were successfully identified in their correct classes using ECG features alone.

On the other hand, the 0.68 macro F1-score confirms that the model maintains a good balance between precision and recall, making it an impressive choice for such cases.

4.5 XGBoost

The confusion matrix of the XGBoost model demonstrates a strong classification performance for CMD detection. It used the final dataset that consists of 97 female subjects and 612 ECG derived features for per subject. The labels were allocated based on PET/CT derived CFR values. If CFR is less than 2.0 , it indicates CMD positive. If CFR is greater than or equal to 2.0, it refers to CMD negative. From Figure 4.5, the model correctly identified 62 CMD negative means no CMD subjects out of 97 subjects and 9 CMD positive. However, 15 CMD negative subjects were incorrectly predicted as CMD positive, and 8 CMD positive subjects were incorrectly identified as no CMD.
This amounts to an overall classification accuracy of 76% which indicates the strong discriminative capability of the model. The recall of XGBoost was 67% which means approximately half of CMD patients were correctly detected. For CMD detection, precision was 0.63, while F1-Score was 0.64. These values are shown in Table 4.1. This reflects that the balance between sensitivity and precision was strong.

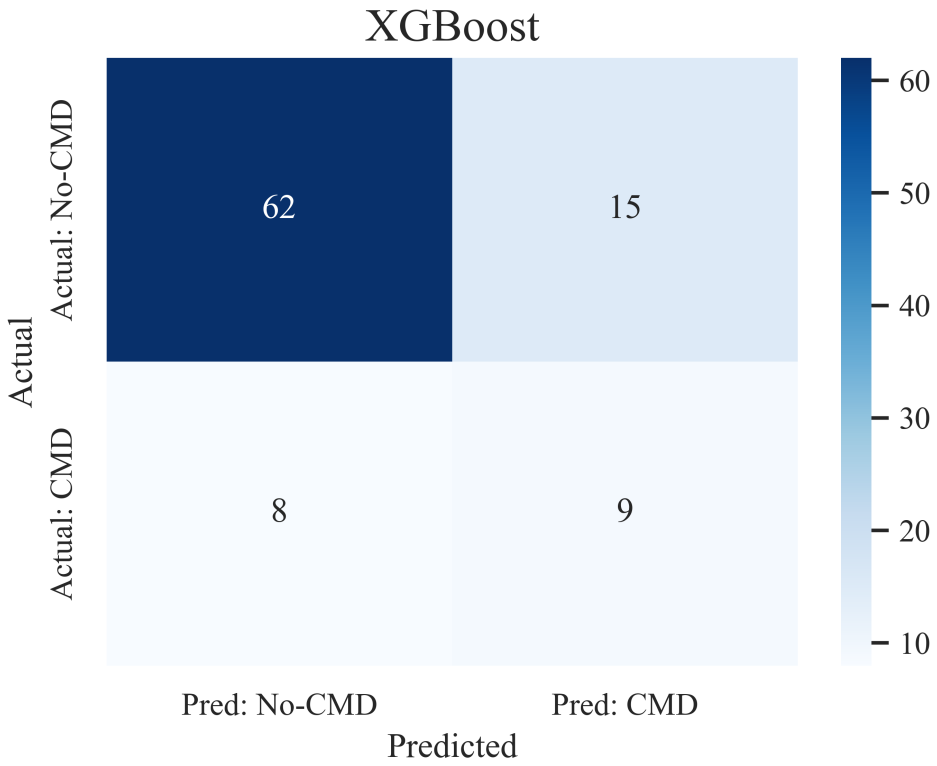


Figure 4.5: Confusion Matrix for XGBoost

Overall, XGBoost performs as strong reliable non-invasive models for CMD detection using ECG based features.

4.6 MLP

From the MLP model, 55 CMD negative subjects were accurately classified as no CMD and 10 CMD positive subjects were correctly detected. On the other hand, 22 CMD negative subjects were incorrectly classified as CMD positive and 7 CMD positive subjects were misclassified as no CMD.

Table 4.1 shows that the accuracy was 69%, recall was 65%, which means approximately three-quarters of true CMD cases were correctly identified. On the other hand, the remaining 26.32% were missed as false negatives. The specificity of the model reached 87.18% and the precision of CMD prediction was around 58.3%. Besides, while the negative predictive value was notably high at approximately 93.2%. The precision was 0.60, and F1-score was 0.60.

Overall, these results shows that the MLP model effectively presents the combined morphological and heart rate variability features across all ECG leads and physiological states to separate the presence of CMD.

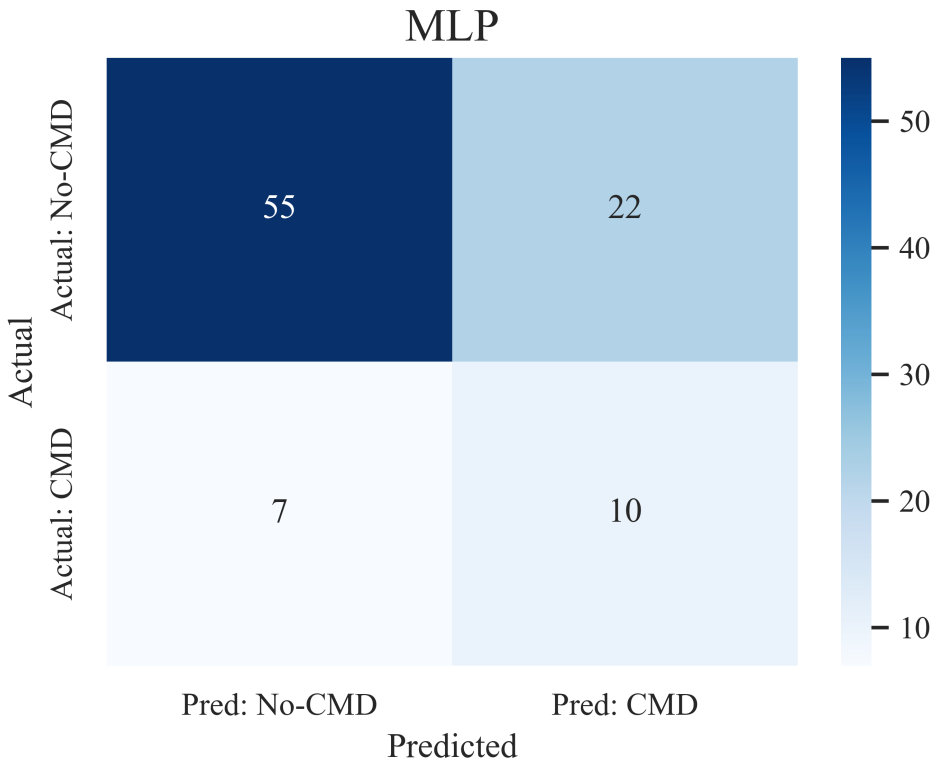


Figure 4.6: Confusion Matrix for MLP

4.7 Ensemble (XGB + RF + SVM)

This model implementing an ensemble of XGBoost, RF and SVMs demonstrated strong performance for detecting CMD using the 612-feature, ECG-derived dataset. The confusion matrix shows that the model correctly classified the major subjects across both classes. From 77 subjects the model accurately predicted 69 subjects as healthy while 8 subjects were incorrectly identified as CMD. For the 17 CMD positive participants, the classifier successfully detected 8 cases and misclassified 9 subject as no CMD. From Table 4.1, it can be seen that the final Ensemble model achieved the highest classification performance with an F1-Score of 0.64, a Recall (Sensitivity) of 67%, and an Accuracy of 82%, precision of 0.63. Combining features from all 12 ECG leads and using information from the 3 stages which are Rest, Stress, and Recovery helps the model make the most reliable predictions for CMD in female patients.

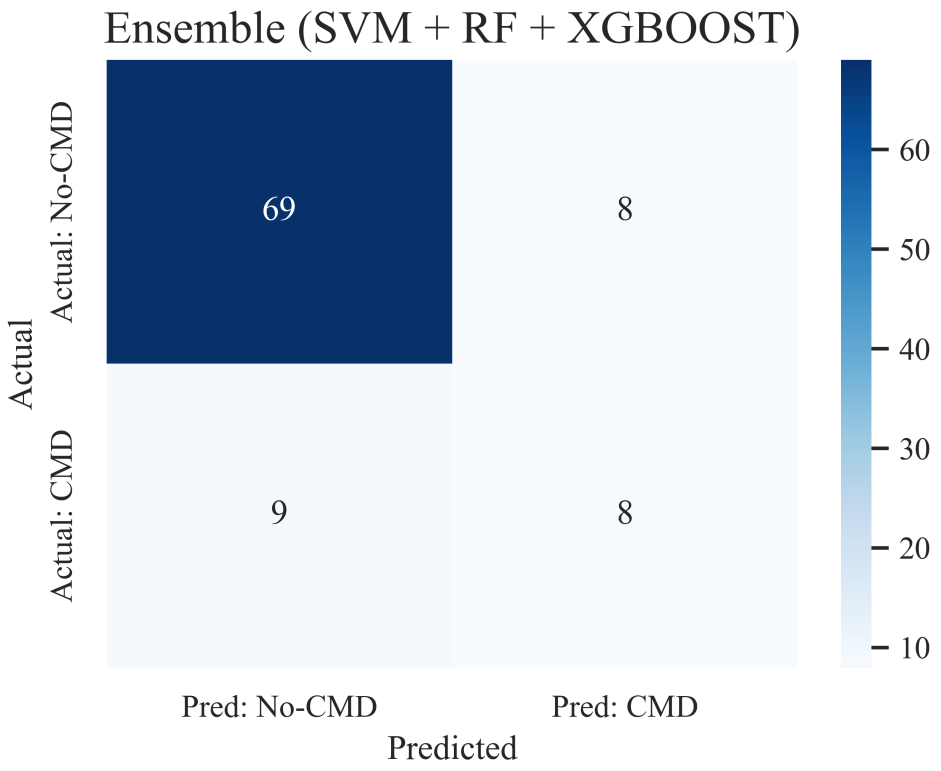


Figure 4.7: Confusion Matrix for Ensemble model

Table 4.1: Performance Comparison of Different Machine Learning Models for CMD Classification

Model	**Accuracy**	**Precision**	**Recall (Sensitivity)**	**F1-Score**
Logistic Regression	64%	0.56	60%	0.55
KNN	64%	0.64	73%	0.60
MLP	69%	0.60	65%	0.60
SVM (RBF)	78%	0.63	63%	0.63
Random Forest	83%	0.71	67%	0.68
XGBoost	76%	0.63	67%	0.64
Ensemble	82%	0.69	68%	0.69

Chapter 5

Discussion

5.1 Objective

The primary objective of this study was to develop an automated machine learning framework for detecting CMD using non-invasive ECG signals. CMD is a difficult to diagnose disease, because unlike obstructive CAD, it affects the small blood vessels of the heart. As a result, many CMD patients show normal ECG at rest, which makes early CMD diagnosis challenging.

Our study demonstrates that CMD can be detected with high accuracy using a carefully designed signal preprocessing pipeline and machine learning based classification. By combining ECG morphological features with HRV features and using ensemble learning, the proposed framework achieved the highest macro F1-score of 0.69. These results confirm that ECG contains the diagnostic patterns related to CMD that are not easily visible to the human eye. This section discusses the comprehensive findings by addressing the research questions and highlighting both the strengths and limitations of the proposed solution.

5.2 The Critical Role of Preprocessing

Our first research question asked how ECG preprocessing affects CMD classification performance. The results of this study clearly show that preprocessing is the most critical step in the entire study.

The ECG data were collected during three physiological phases: Rest, Adenosine-induced Stress, and Recovery. Although adenosine is effective for simulating cardiac stress, the ECG signals contained significant noise across all phases, not only during stress, which is a major problem for signal processing. Patients breathe very heavily, and their chests move rapidly. Artifacts or voltage spikes can be caused by multiple factors at any stage. It includes patient movement, electrode displacement, heavy breathing,

anxiety, and muscle activity. These disturbances are the reason for baseline wander, sharp voltage spikes, and irregular signal distortions, and the computer confused these noisy spikes with heart disease markers [21].

Initial experiments using standard filtering techniques, such as a simple band-pass filter, were unsuccessful. The models misinterpreted noise artifacts as disease-related signal changes, leading to poor classification performance. This proved that conventional preprocessing was not sufficient for this dataset. To address this issue, we designed a dual-path pipeline.

- **Detection Path:** Aggressive cleaning techniques were applied, including wavelet denoising with Daubechies-6, to sharpen the ECG signal and enhance R-peak detection. This allowed us to find beats even during heavy noise.
- **Measurement Path:** A gentle Butterworth band-pass filter was applied to preserve the original morphology of the ECG waveform. This was especially important for maintaining the original height of the ST-segment and T-wave, which are important for ischemia detection.

In addition, a template matching technique was used. For each subject, an average heartbeat template was created. Each detected beat was compared against this template. The beats with low correlation were classified as artifacts and removed. This strategy significantly improved data reliability in medical AI where data quality is more important than model complexity.

5.3 Feature Importance and Physiological Markers

The second research question aimed to identify which ECG features are most informative for CMD detection.

5.3.1 The Value of "Delta" Features

The study found that absolute ECG values during the rest phase were often not sufficient for CMD detection. This is expected, as many CMD patients show normal ECG patterns when the heart is not under stress.

Delta features represent the mathematical difference between the Stress and Rest phases. These features capture how the heart responds to increased oxygen demand. Since CMD is a functional disorder where microvessels fail to dilate properly during stress, Delta features effectively represent the disease mechanism. By teaching the model to look at the change (Delta), we captured the heart's inability to adapt to stress.

5.3.2 Morphology vs. HRV

Feature selection based on Mutual Information revealed the importance of two feature categories:

- **Morphological (Ischemia) Features:** Features like mean ST-segment and T-wave amplitude were critical. These represent the electrical recovery of the heart muscle, which is known to be affected by lack of oxygen (ischemia).
- **HRV (Autonomic) Features:** Features like RMSSD and SD1 were also selected. These features reflect autonomic nervous system regulation of the heart. Their relevance suggests that CMD is not only a vascular disorder but also affects autonomic cardiac control, the nerves that control the heart.

The presence of both morphology-based and HRV-based features confirms that a multi-feature ECG representation is more powerful than relying on a single category of markers.

5.4 Model Performance and Class Imbalance

The final dataset consisted of 94 patient data, including only 17 CMD patients and 77 healthy subjects. This class imbalance created a significant challenge for model training. Our third research question dealt with these challenges.

5.4.1 Failure of Simple Models

Linear models, such as LR, showed high accuracy but extremely poor macro Recall (around 60%). This occurred because the model predicted most subjects as healthy, which is statistically true when the majority class (non-CMD) dominates the dataset. However, missing diseased patients is not acceptable in clinical applications. This highlights why accuracy alone is a misleading metric in imbalanced medical datasets. This limitation became more evident when macro precision, macro recall, and macro F1-score were considered, as these metrics give equal importance to both CMD and non-CMD classes.

5.4.2 Success of the Ensemble Model

Tree-based models such as RF and XGBoost performed significantly better because they are capable of modeling complex non-linear relationships. While Random Forest achieved the highest overall accuracy of 83%, the Ensemble model showed more balanced performance for CMD detection, maintaining the second accuracy of 82%. The Ensemble model, which combined SVM, RF, and XGBoost used the soft voting method to calculate the results. By averaging probability scores, the ensemble balanced individual model

weaknesses and achieved a macro Recall of 68%, ensuring the balance between predicting both CMD and non-CMD classes. This highlights that, in imbalanced medical datasets, model evaluation should prioritize macro-averaged metrics and recall rather than accuracy to ensure reliable disease classification.

To further ensure reliability, LOSO cross-validation was used. In this strict testing scheme, the model was trained on 93 subjects and tested on the remaining one subject, repeating the process 94 times. This ensures that our high accuracy is real and not just a result of the model "memorizing" the patients.

5.5 Limitations of the Study

Despite the promising results, several limitations of the study should be acknowledged.

- **Limited Sample Size:** The final dataset contains only 94 subjects, which is relatively small, especially for deep learning models. Although LOSO validation was used, external validation on independent clinical populations is required to confirm the generalization of the proposed method.
- **Adenosine vs. Exercise Stress:** Adenosine mainly alters coronary blood flow, whereas physical exercise also increases heart rate and blood pressure. Since the dataset is based on adenosine stress, the trained model may not generalize equally well to treadmill-based ECG stress testing.
- **Removal of Artifacts:** While removing noisy beats improved signal quality, excessive artifact rejection may reduce the amount of usable data for some patients. In a real-world clinical setting, if a patient moves too much, the system might reject too many beats and fail to give a diagnosis.
- **Black-Box Nature of the Model:** Although we know which features were selected, the Ensemble model is complex. It is difficult to explain exactly how the model combines a T-wave change with an HRV change to make a specific decision. This "Black Box" issue can make it harder for doctors to trust the system completely.

Chapter 6

Conclusion and Future Work

6.1 Summary of the Study

CMD is one of the most difficult heart conditions to diagnose. Unlike obstructive CAD, which shows visible blockages, CMD originates from dysfunction in the small coronary vessels and therefore does not appear in standard imaging. As a result, diagnosis often requires invasive and expensive functional tests such as PET-based coronary flow reserve.

The primary goal of this thesis was to explore whether a fully automated and non-invasive screening system could be developed using standard 12-lead ECG signals combined with machine learning techniques.

The study used a dataset of 94 women (17 CMD patients and 77 healthy controls) who underwent an adenosine stress protocol. A major challenge was the high level of noise during the stress phase caused by breathing and body movement, which obscured the subtle physiological information needed for CMD detection.

To address this, a "Dual-Path" preprocessing strategy was designed. This approach separated heartbeat detection from waveform measurement, allowing each task to be optimized independently. After cleaning, the ECG signals were used to extract both waveform morphology features and HRV metrics. A feature selection process was applied, and the final classifier combined SVM, RF and XGBoost in a soft voting ensemble architecture.

6.2 Key Conclusions

From the findings of this study, several key conclusions were drawn:

- **Preprocessing is the Key to Success:** Standard filtering methods are not sufficient for Adenosine stress ECGs. The "Electrode Pop" artifacts caused by heavy breathing can ruin the signals. Template match method was used to detect and

remove these bad beats which was very essential. Without this step, the machine learning models could not perform effectively.

- **Dynamic Changes Matter Most:** The most important features were not found from the "rest" phase. The most important features came from the "delta" features of the difference between the "stress" phase and the "rest" phase. CMD is a dynamic condition where the heart looks normal at rest but fails under stress.
- **Multi-Feature Approach is Necessary:** The best performance came from combining ischemia related features such as ST-segment changes with autonomic markers such as HRV. This suggests that using only one type of feature was not enough to distinguish CMD subject from healthy subject.
- **Ensemble Model Handles Imbalance Effectively:** Despite having only 17 CMD cases, the ensemble model evaluated using LOSO cross validation achieved a macro sensitivity of 68% and a macro F1-score of 0.69. This demonstrates the robustness of ensemble approaches for imbalanced medical datasets.

6.3 Contributions to the Field

This study contributes to biomedical engineering and computational cardiology in the following ways:

- It introduces a reproducible preprocessing pipeline tailored to the noise characteristics of pharmacological stress ECGs.
- It identifies and ranks twelve key ECG features associated with microvascular dysfunction, guiding future investigations.
- It provides evidence that, automated and non-invasive CMD detection using standard ECG signals is feasible and can show stronger performance, potentially reducing reliance on invasive initial screening tests.

6.4 Future Work

Although this study provides a strong foundation, several directions remain open for future research:

6.4.1 Improving Explainability and Clinical Trust

This study focused on accuracy, but in real hospitals, doctors need to understand why the AI made a decision. Future research should use Explainable AI (XAI) tools such as

SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-Agnostic Explanations) to create visual reports that show which parts of the ECG influenced the CMD prediction.

6.4.2 Expanding to Deep Learning with Larger Datasets

The dataset size has limited model complexity. With larger multi-centre datasets, future work could explore deep learning models such as CNNs or LSTMs that may identify subtle signal patterns beyond the reach of handcrafted features.

6.4.3 External and Multi-Centre Validation

To ensure clinical relevance, the model must demonstrate robust performance across diverse populations, ECG devices, and stress protocols (both pharmacological and exercise-based). Such validation is essential for real-world generalisability.

6.4.4 Real-Time Implementation

Currently, the analysis is done offline. Future work could focus on creating a real-time system that shows the CMD risk score during the stress test, providing immediate feedback to technicians and doctors.

6.5 Final Remarks

This thesis demonstrates that although CMD produces subtle electrical signals, they can be detected through careful preprocessing and machine learning analysis. By emphasizing the heart's dynamic response to pharmacological stress and treating noise reduction as a central scientific challenge, this work extracts meaningful diagnostic information from standard ECG recordings.

This study demonstrates that hidden signs of CMD can be detected using signal processing and machine learning models. By carefully handling noise and focusing on the heart's dynamic response to stress, standard ECG signals can be turned into a powerful diagnostic tool. This work lays the foundation for future technologies that make heart screening more accurate, accessible, and non-invasive for everyone.

Appendix

A.1 Appendix A: Abbreviations

1. CMD: Coronary Microvascular Disease
2. CAD: Coronary Artery Disease
3. CVD: Cardiovascular Disease
4. ECG: Electrocardiogram
5. HRV: Heart Rate Variability
6. CFR: Coronary Flow Reserve
7. PET/CT: Positron Emission Tomography / Computed Tomography
8. ANOCA: Angina with non-Obstructive Coronary Arteries
9. INOCA: Ischemia with non-Obstructive Coronary Arteries
10. ML: Machine Learning
11. DL: Deep Learning
12. SVM: Support Vector Machine
13. RF: Random Forest
14. KNN: K-Nearest Neighbors
15. MLP: Multilayer Perceptron
16. MI: Mutual Information
17. LOSO: Leave-One-Subject-Out
18. RBF: Radial Basis Function
19. SD: Standard Deviation

20. CV: Cross validation
21. LR: Logistic Regression
22. SDG: Sustainable Development Goal
23. SHAP: Shapley Additive Explanations
24. LIME: Local Interpretable Model-Agnostic Explanations

Bibliography

[1] World Health Organization. Cardiovascular diseases (cvds) — fact sheet. https://www.who.int/health-topics/cardiovascular-diseases, 2024. Accessed: 2025-11-07.

[2] Centers for Disease Control and Prevention (CDC). About heart disease. https://www.cdc.gov/heart-disease/about/index.html, 2024. Accessed: 2025-11-07.

[3] Viviany R Taqueti and Marcelo F Di Carli. Coronary microvascular disease pathogenic mechanisms and therapeutic options: Jacc state-of-the-art review. *Journal of the American College of Cardiology*, 72(21):2625–2641, 2018.

[4] Sujith Kuruvilla and Christopher M Kramer. Coronary microvascular dysfunction in women: an overview of diagnostic strategies. *Expert review of cardiovascular therapy*, 11(11):1515–1525, 2013.

[5] Venkatesh L Murthy, Masanao Naya, Viviany R Taqueti, Courtney R Foster, Mariya Gaber, Jon Hainer, Sharmila Dorbala, Ron Blankstein, Ornella Rimoldi, Paolo G Camici, et al. Effects of sex on coronary microvascular dysfunction and cardiac outcomes. *Circulation*, 129(24):2518–2527, 2014.

[6] A. Sinha, H. Rahman, and D. Perera. Coronary microvascular disease: current concepts of pathophysiology, diagnosis and management. *Cardiovascular Endocrinology and Metabolism*, 10:22–30, 2021.

[7] T. J. Ford, B. Stanley, R. Good, P. Rocchiccioli, M. McEntegart, S. Watkins, et al. Stratified medical therapy using invasive coronary function testing in angina: the cormica trial. *Journal of the American College of Cardiology*, 72(23):2841–2855, 2018.

[8] Viviany R. Taqueti and Marcelo F. Di Carli. Coronary microvascular disease pathogenic mechanisms and therapeutic options. *JACC*, 72(21):2625–2641, 2018.

[9] Benjamin J.W. Chow, Sharmila Dorbala, Marcelo F. Di Carli, Michael E. Merhige, Brent A. Williams, Emir Veledar, James K. Min, Michael J. Pencina, Yeung Yam, Li Chen, Sai Priya Anand, Terrence D. Ruddy, Daniel S. Berman, Leslee J. Shaw, and

Rob S. Beanlands. Prognostic value of pet myocardial perfusion imaging in obese patients. *JACC: Cardiovascular Imaging*, 7(3):278–287, 2014.

[10] Fabio Mangiacapra, Michele Mattia Viscusi, Giuseppe Verolino, Luca Paolucci, Annunziata Nusca, Rosetta Melfi, Gian Paolo Ussia, and Francesco Grigioni. Invasive assessment of coronary microvascular function. *Journal of Clinical Medicine*, 11(1):228, 2021.

[11] Vijay Kunadian, Alaide Chieffo, Paolo G Camici, Colin Berry, Javier Escaned, Angela HEM Maas, Eva Prescott, Nicole Karam, Yolande Appelman, Chiara Fraccaro, et al. An eapci expert consensus document on ischaemia with non-obstructive coronary arteries in collaboration with european society of cardiology working group on coronary pathophysiology & microcirculation endorsed by coronary vasomotor disorders international study group. *European heart journal*, 41(37):3504–3520, 2020.

[12] Thomas J Ford, Bethany Stanley, Richard Good, Paul Rocchiccioli, Margaret McEntegart, Stuart Watkins, Hany Eteiba, Aadil Shaukat, Mitchell Lindsay, Keith Robertson, et al. Stratified medical therapy using invasive coronary function testing in angina: the cormica trial. *Journal of the American College of Cardiology*, 72(23 Part A):2841–2855, 2018.

[13] Laura Viola, Megan Masters, Umar Shafiq, Krishnam Raju Jujjavarapu, and Suvitesh Luthra. Ischemia with no obstructive coronary artery disease (inoca): A review. *Life*, 15(10):1554, 2025.

[14] Islam Y Elgendy, Lina Ya'Qoub, Kuan-Han Chen, and Carl J Pepine. Coronary microvascular dysfunction in patients with non-obstructive coronary arteries: current gaps and future directions. *Drugs*, 82(3):241–250, 2022.

[15] Aish Sinha, Haseeb Rahman, and Divaka Perera. Coronary microvascular disease: current concepts of pathophysiology, diagnosis and management. *Cardiovascular endocrinology & metabolism*, 10(1):22–30, 2021.

[16] Viviany R Taqueti. Coronary microvascular dysfunction in heart failure with preserved ejection fraction—common, unrecognized, and prevalent in patients with or without epicardial cad. *JAMA cardiology*, 6(10):1118–1120, 2021.

[17] Alejandro Travieso, Adrian Jeronimo-Baza, Daniel Faria, Asad Shabbir, Hernan Mejia-Renteria, and Javier Escaned. Invasive evaluation of coronary microvascular dysfunction. *Journal of Nuclear Cardiology*, 29(5):2474–2486, 2022.

[18] Ravi A Thakker, Jorge Rodriguez Lozano, Patricia Rodriguez Lozano, Afaq Motiwala, Umamahesh Rangasetty, Wissam Khalife, and Khaled Chatila. Coronary microvascular disease. *Cardiology and Therapy*, 11(1):23–31, 2022.

[19] Eva Prescott, Kira Bang Bove, Daria Frestad Bechsgaard, Bilal Hasan Shafi, Theis Lange, Jakob Schroder, Hanna Elena Suhrs, and Rikke Linnemann Nielsen. Biomarkers and coronary microvascular dysfunction in women with angina and no obstructive coronary artery disease. *JACC: advances*, 2(2):100264, 2023.

[20] Shubhojeet Chatterjee, Rini Smita Thakur, Ram Narayan Yadav, Lalita Gupta, and Deepak Kumar Raghuvanshi. Review of noise removal techniques in ecg signals. *IET Signal Processing*, 14(9):569–590, 2020.

[21] Xiaoting Peng and Dantong Li. A multimodal dataset to explore the potential biomarker for coronary microvascular disease, April 2025.

[22] AHM Waliul Islam, AQM Reza, Shams Munwar, Shahabuddin Talukder, Tamzeed Ahmed, Kazi Atikur Rahman, and Md Atahar Ali. Ischemia of non-obstructive coronary artery disease: Need to address in our patient population. *Cardiovascular Journal*, 13(2):217–222, 2021.

[23] Juhani Knuuti, William Wijns, Antti Saraste, Davide Capodanno, Emanuele Barbato, Christian Funck-Brentano, Eva Prescott, Robert F Storey, Christi Deaton, Thomas Cuisset, et al. 2019 esc guidelines for the diagnosis and management of chronic coronary syndromes: The task force for the diagnosis and management of chronic coronary syndromes of the european society of cardiology (esc). *European heart journal*, 41(3):407–477, 2020.

[24] THE 17 GOALS | Sustainable Development — sdgs.un.org. `https://sdgs.un.org/goals`. [Accessed 05-12-2025].

[25] Rahul Kumar Sevakula, Wan-Tai M Au-Yeung, Jagmeet P Singh, E Kevin Heist, Eric M Isselbacher, and Antonis A Armoundas. State-of-the-art machine learning techniques aiming to improve patient outcomes pertaining to the cardiovascular system. *Journal of the American Heart Association*, 9(4):e013924, 2020.

[26] Fatemeh Zabihollahy, S Rajan, and E Ukwatta. Machine learning-based segmentation of left ventricular myocardial fibrosis from magnetic resonance imaging. *Current Cardiology Reports*, 22(8):65, 2020.

[27] M Lalli and S Amutha. Applications of deep learning and machine learning in healthcare domain-a literature review. *International Journal of Electrical Engineering and Technology*, 11(8):113–126, 2020.

[28] Albert K Feeny, Mina K Chung, Anant Madabhushi, Zachi I Attia, Suraj, Sanjiv M Narayan, et al. Artificial intelligence and machine learning in arrhythmias and cardiac electrophysiology. *Circulation: Arrhythmia and Electrophysiology*, 13(8):e007952, 2020.

[29] Partha Sardar, J Dawn Abbott, Amartya Kundu, Herbert D Aronow, Juan F Granada, and Jay Giri. Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance. *Cardiovascular interventions*, 12(14):1293–1303, 2019.

[30] Yogesh NV Reddy, Masaru Obokata, Frederik H Verbrugge, Grace Lin, and Barry A Borlaug. Atrial dysfunction in patients with heart failure with preserved ejection fraction and atrial fibrillation. *Journal of the American College of Cardiology*, 76(9):1051–1064, 2020.

[31] Tianyi Liu, Andrew Krentz, Lei Lu, and Vasa Curcin. Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis. *European Heart Journal-Digital Health*, 6(1):7–22, 2025.

[32] Samah J Fodeh, Taihua Li, Haya Jarad, and Basmah Safdar. Classification of patients with coronary microvascular dysfunction. *IEEE/ACM transactions on computational biology and bioinformatics*, 17(2):704–711, 2019.

[33] Priya Koilpillai, Niti R Aggarwal, and Sharon L Mulvagh. State of the art in noninvasive imaging of ischemic heart disease and coronary microvascular dysfunction in women: indications, performance, and limitations. *Current Atherosclerosis Reports*, 22(12):73, 2020.

[34] Eleni I Georga, Nikolaos S Tachos, Antonis I Sakellarios, Gualtiero Pelosi, Silvia Rocchiccioli, Oberdan Parodi, Lampros K Michalis, and Dimitrios I Fotiadis. A multimodal machine learning approach to omics-based risk stratification in coronary artery disease. In *World Congress on Medical Physics and Biomedical Engineering 2018: June 3-8, 2018, Prague, Czech Republic (Vol. 3)*, pages 879–882. Springer, 2018.

[35] Ahmed E Fetit, Alexander S Doney, Stephen Hogg, Ruixuan Wang, Tom MacGillivray, Joanna M Wardlaw, Fergus N Doubal, Gareth J McKay, Stephen McKenna, and Emanuele Trucco. A multimodal approach to cardiovascular risk stratification in patients with type 2 diabetes incorporating retinal, genomic and clinical features. *Scientific reports*, 9(1):3591, 2019.

[36] T. Liu, P. Li, Y. Liu, H. Zhang, Y. Li, Y. Jiao, C. Liu, C. Karmakar, X. Liang, M. Ren, and X. Wang. Detection of coronary artery disease using multi-domain feature fusion of multi-channel heart sound signals. *Entropy (Basel, Switzerland)*, 23(6):642, 2021.

[37] M. Pal, S. Parija, G. Panda, K. Dhama, and R. K. Mohapatra. Risk prediction of cardiovascular disease using machine learning classifiers. *Open Medicine (Warsaw, Poland)*, 17(1):1100–1113, 2022.

[38] C. A. U. Hassan, J. Iqbal, R. Irfan, S. Hussain, A. D. Algarni, S. S. H. Bukhari, N. Alturki, and S. S. Ullah. Effectively predicting the presence of coronary heart disease using machine learning classifiers. *Sensors (Basel, Switzerland)*, 22(19):7227, 2022.

[39] M. Sayadi, V. Varadarajan, F. Sadoughi, S. Chopannejad, and M. Langarizadeh. A machine learning model for detection of coronary artery disease using noninvasive clinical parameters. *Life*, 12(11):1933, 2022.

[40] Kanyifeechukwu Oguine, Ozioma Oguine, and Chito Ofodum. Detection of significant risk factors of heart diseases by educing multimodal features and implementing machine learning techniques. *International Journal of Scientific Research in Science and Technology*, pages 621–630, 08 2021.

[41] E. Panteris, O. Deda, A. S. Papazoglou, E. Karagiannidis, T. Liapikos, O. Begou, T. Meikopoulos, T. Mouskeftara, G. Sofidis, G. Sianos, G. Theodoridis, and H. Gika. Machine learning algorithm to predict obstructive coronary artery disease: Insights from the corlipid trial. *Metabolites*, 12(9):816, 2022.

[42] A. Banerjee, S. Chen, G. Fatemifar, M. Zeina, R. T. Lumbers, J. Mielke, S. Gill, D. Kotecha, D. F. Freitag, S. Denaxas, and H. Hemingway. Machine learning for subtype definition and risk prediction in heart failure, acute coronary syndromes and atrial fibrillation: Systematic review of validity and clinical utility. *BMC Medicine*, 19(1):85, 2021.

[43] G. Quer, R. Arnaout, M. Henne, and R. Arnaout. Machine learning and the future of cardiovascular care: Jacc state-of-the-art review. *Journal of the American College of Cardiology*, 77(3):300–313, 2021.

[44] S. Amal, L. Safarnejad, J. A. Omiye, I. Ghanzouri, J. H. Cabot, and E. G. Ross. Use of multi-modal data and machine learning to improve cardiovascular disease care. *Frontiers in Cardiovascular Medicine*, 9:840262, 2022.

[45] A. Baggiano, G. Italiano, M. Guglielmo, L. Fusini, A. I. Guaricci, R. Maragna, C. M. Giacari, S. Mushtaq, E. Conte, A. D. Annoni, A. Formenti, M. E. Mancini, D. Andreini, M. Rabbat, M. Pepi, and G. Pontone. Changing paradigms in the diagnosis of ischemic heart disease by multimodality imaging. *Journal of Clinical Medicine*, 11(3):477, 2022.

[46] Rubini Pe, C Subasini, A Katharine, V Kumaresan, S Gowdhamkumar, and T Nithya. A cardiovascular disease prediction using machine learning algorithms. *Annals of the Romanian Society for Cell Biology*, 25:904–912, 01 2021.

[47] Shintaro Fujisawa and Yusuke Fukazawa. Multimodal transformer for heart disease classification using multiple heart sound spectral analyses and clinical metadata. 2025.

[48] Xin Zhang, Liangxiu Han, Stephen White, Saad Hassan, Philip A Kalra, James Ritchie, Carl Diver, and Jennie Shorley. Tabulatime: A novel multimodal deep learning framework for advancing acute coronary syndrome prediction through environmental and clinical data integration. *arXiv preprint arXiv:2502.17049*, 2025.

[49] Chenglin Lin, Huimin Lu, Pengcheng Sang, and Chenyu Pan. A knowledge embedded multimodal pseudo-siamese model for atrial fibrillation detection. *Scientific Reports*, 15(1):3133, 2025.

[50] Ghadah Alwakid, Farman Ul Haq, Noshina Tariq, Mamoona Humayun, Momina Shaheen, and Marwa Alsadun. Optimized machine learning framework for cardiovascular disease diagnosis: a novel ethical perspective. *BMC Cardiovascular Disorders*, 25(1):123, 2025.

[51] Ahmed Alsayat, Alshimaa Abdelraof Mahmoud, Saad Alanazi, Ayman Mohamed Mostafa, Nasser Alshammari, Majed Abdullah Alrowaily, Hosameldeen Shabana, and Mohamed Ezz. Enhancing cardiac diagnostics: A deep learning ensemble approach for precise ecg image classification. *Journal of Big Data*, 12(1):7, 2025.

[52] Ashish Shiwlani, Ahsan Ahmad, Muhammad Umar, Nasrullah Dharejo, Anoosha Tahir, and Sheena Shiwlani. Analysis of multi-modal data through deep learning techniques to diagnose cvds: A review. *International Journal*, 11(1):402–420, 2024.

[53] Victor Adeyi, Zhang Xiaoling, Kuyikabasi Uko, Izis Kanjarawy, and Gyarteng Emmanuel. Cardionet+: Revolutionizing heart failure diagnosis with multi-modal learning. *Available at SSRN 5132486*, 2025.

[54] Junxin Wang, Juanen Li, Rui Wang, and Xinqi Zhou. Vae-driven multimodal fusion for early cardiac disease detection. *IEEE Access*, 12:90535–90551, 2024.

[55] Muhammet Alkan, Gruschen Veldtman, and Fani Deligianni. Predicting cardiopulmonary exercise testing outcomes in congenital heart disease through multi-modal data integration and geometric learning. *arXiv preprint arXiv:2503.14239*, 2025.

[56] Ran Xiao, Cheng Ding, Xiao Hu, Gari D Clifford, David W Wright, Amit J Shah, Salah Al-Zaiti, and Jessica K Zègre-Hemsey. Integrating multimodal information in

machine learning for classifying acute myocardial infarction. *Physiological Measurement*, 44(4):044002, 2023.

[57] Bo Yu, Hechang Chen, Chengyou Jia, Hongren Zhou, Lele Cong, Xiankai Li, Jianhui Zhuang, and Xianling Cong. Multi-modality multi-scale cardiovascular disease subtypes classification using raman image and medical history. *Expert Systems with Applications*, 224:119965, 2023.

[58] Jiajian Zhu, Yue Feng, Qichao Liu, Hong Xu, Yuan Miao, Zhuosheng Lin, Jia Li, Huilin Liu, Ying Xu, and Fufeng Li. An improved convnext with multimodal transformer for physiological signal classification. *IEEE Access*, 12:11217–11229, 2024.

[59] Dantong Li, Xiaoting Peng, Lianting Hu, Jintai Chen, Xinyang Long, Xueli Zhang, Siting Ye, Xiaohe Bai, Chao Wu, Huan Yang, et al. A multimodal dataset for coronary microvascular disease biomarker discovery. *Scientific Data*, 12(1):990, 2025.

[60] Ana Margarida Martins, Miguel Nobre Menezes, Pedro Alves da Silva, and Ana G Almeida. Multimodality imaging in the diagnosis of coronary microvascular disease: An update. *Journal of Personalized Medicine*, 15(2):75, 2025.

[61] Kaushalya Dissanayake and Md Gapar Md Johar. Comparative study on heart disease prediction using feature selection techniques on classification algorithms. *Applied Computational Intelligence and Soft Computing*, 2021:1–17, 11 2021.

[62] David Moher, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G Altman. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Bmj*, 339, 2009.

[63] Coronary Flow Reserve (CFR) — my.clevelandclinic.org. `https://my.clevelandclinic.org/health/diagnostics/24028-coronary-flow-reserve`. [Accessed 05-12-2025].

[64] What Is an Adenosine Stress Test? — my.clevelandclinic.org. `https://my.clevelandclinic.org/health/diagnostics/16839-cardiac-mri-adenosine-stress-test`. [Accessed 05-12-2025].

[65] Dominique Makowski, Tam Pham, Zen J Lau, Jan C Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and SH Annabel Chen. Neurokit2: A python toolbox for neurophysiological signal processing. *Behavior research methods*, 53(4):1689–1696, 2021.

[66] Park Hyeoun-Ae. An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *jkan*, 43(2):154–164, 2013.

[67] Javier M. Moguerza and Alberto Muñoz. Support Vector Machines with Applications. *Statistical Science*, 21(3):322 – 336, 2006.

[68] Hasan Salman, Ali Kalakech, and Amani Steiti. Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024:69–79, 06 2024.

[69] Kashvi Taunk, Sanjukta De, Srishti Verma, and Aleena Swetapadma. A brief review of nearest neighbor algorithm for learning and classification. pages 1255–1260, 2019.

[70] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. pages 785–794, 08 2016.

[71] Kit Yan Chan, Bilal Abu-Salih, Raneem Qaddoura, Ala'M Al-Zoubi, Vasile Palade, Duc-Son Pham, Javier Del Ser, and Khan Muhammad. Deep neural networks in the cloud: Review, applications, challenges and research directions. *Neurocomputing*, 545:126327, 2023.

[72] Mohammad Hossin and Sulaiman M.N. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining Knowledge Management Process*, 5:01–11, 03 2015.

[73] Understanding the Confusion Matrix in Machine Learning - GeeksforGeeks — geeksforgeeks.org. `https://www.geeksforgeeks.org/machine-learning/confusion-matrix-machine-learning/`. [Accessed 20-11-2025].