DEVELOPMENT OF A MACHINE LEARNING MODEL FOR AUTOMATED DIAGNOSIS OF ECG ABNORMALITIES

Raiymbek Nurtay, Bachelor of Science

Submitted in fulfillment of the requirements for the degree of a Master of Science in Biomedical Engineering

School of Engineering and Digital Sciences

Department of Chemical and Materials Engineering

Nazarbayev University

53 Kabanbay Batyr Avenue,

Astana, Kazakhstan, 010000

Supervisors:

Lead Supervisor Name: Associate Professor Cevat Erisken

Co-supervisor Name: Professor Adnan Yazici

April 2025

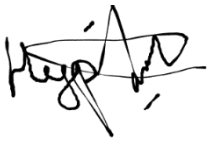
Declaration

I hereby declare that this manuscript, entitled “Development of a Machine Learning Model for Automated Diagnosis of ECG Abnormalities” is the result of my work except for quotations and citations which have been duly acknowledged.

I also declare that, to the best of my knowledge and belief, it has not been previously or concurrently submitted, in whole or in part, for any other degree or diploma at Nazarbayev University or any other national or international institution.

This research has been conducted with the approval of the Nazarbayev University Institutional Research Ethics Committee (NU-IREC) under the approval number 1033/05032025.

Name: Raiymbek Nurtay



Date: 04.04.2025

Acknowledgements

I send my most heartfelt appreciation to my family who continuously supported me during every step of my academic education. Their faith and love for me has proven itself a steady well of power throughout my life.

Professor Cevat Erisken has been instrumental in my growth through his direction along with his precious academic knowledge and his persistent motivation. His expert guidance along with mentoring duties played an essential role in developing my research progress and academic development. I express appreciation to Professor Adnan Yazici for his regular support together with his thoughtful recommendations and valuable work contributions towards this research outcome.

I am thankful to all my friends because their motivating support enables me to grow. Their dual academic and personal backing has transformed this journey into a rewarding and more pleasant experience.

I express appreciation to everyone who helped in finishing this thesis through direct or indirect involvement. The completion of this work required the assistance along with support from numerous people.

Table of Contents

List of abbreviations

CVD	Cardiovascular Disease
ECG	Electrocardiogram
IHD	Ischemic Heart Disease
AV	Atrioventricular
SA	Sinoatrial
PCA	Principal Component Analysis
DWT	Discrete Wavelet Transform
SVM	Support Vector Machine
k-NN	k-Nearest Neighbors
Naïve Bayes	Naive Bayes Classifier
LDA	Linear Discriminant Analysis
CART	Classification and Regression Trees
CNN	Convolutional Neural Network
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
F1-Score	F1 Weighted Score
PR	Precision-Recall
CI	Confidence Interval
SNR	Signal-to-Noise Ratio
P-wave	Atrial Depolarization Wave
QRS	Ventricular Depolarization Complex

T-wave	Ventricular Repolarization Wave
ST-segment	Segment between S and T waves
MI	Myocardial Infarction
STTC	ST-T Change
CD	Conduction Disturbance
HYP	Hypertrophy
REM	Resampling
REMEDIAL	Resampling Multi-label Datasets by Decoupling Highly Imbalanced Labels
MLeNN	Multi-label Edited Nearest Neighbors
IRLbl	Imbalance Ratio per Label
SCUMBLE	Score of Co-occurrence among Imbalanced Labels
Jaccard Index	Similarity coefficient between sets
GPU	Graphics Processing Unit
CI/CD	Continuous Integration/Continuous Deployment
UI	User Interface
UX	User Experience
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
F1	F1 Score

TPR	True Positive Rate
TNR	True Negative Rate
FPR	False Positive Rate
FNR	False Negative Rate

List of Figures and Tables

Abstract

Early diagnosis of Cardiovascular diseases (CVD) remains essential because these diseases represent the main global reason for mortality and disease incidents. Traditional ECG analysis methods struggle to identify minor and intricate abnormalities within extensive databases because Electrocardiograms (ECGs) serve as standard tools to inspect heart health conditions. The research creates a hybrid machine learning model for automated ECG abnormality detection which integrates deep learning features of CNN (1D Convolutional Neural Network) for feature extraction with classifier chain ensemble modeling for multi-label classification. The experimental results based on the PTB-XL database demonstrated an overall accuracy rate of 87.79% together with a weighted accuracy of 95.98% through model training and evaluation. This performance exceeded standard methods. The main contribution of this study includes connecting a machine learning algorithm to a web-based interface which enables healthcare professionals especially those in limited resource environments to submit single-lead ECG images that produce instant diagnostic results. This integration facilitates faster, more efficient ECG analysis with a response time of approximately 22 seconds per image. The hybrid system represents progress in medical diagnostic workstreams and expands automatic cardiovascular disease detection through its monitoring platform despite facing difficulties in detecting uncommon abnormalities. Studies prove that machine learning-based ECG diagnostic systems can successfully enter clinical settings where they create a practical and quick method to enhance cardiovascular treatment.

Keywords: *Cardiovascular Disease, Electrocardiogram, Machine Learning, Hybrid Model, Remote Healthcare.*

1 CHAPTER 1- (INTRODUCTION)

1.1 Cardiovascular Disease

Cardiovascular disease (CVD) encompasses multiple disorders which affect the heart and blood vessels including ischemic heart disease (IHD) stroke heart failure and peripheral artery disease. The conditions emerge from atherosclerosis which leads to fatty deposits building up in arterial walls to block blood flow. CVD development results from a combination of changeable and unchangeable risk factors which include hypertension and diabetes along with obesity and smoking and a lack of physical activity.

1.1.1 Global Mortality and Socioeconomic Disparities in Cardiovascular Disease

The early identification and management of CVD risk stands essential to decrease mortality and morbidity because cardiovascular disease causes 17.5 million worldwide deaths annually [1].

The death rates from cardiovascular diseases have decreased in high-income nations but Eastern European countries face elevated mortality rates compared to other low-income regions. The observed differences in CVD outcomes between populations demonstrate how economic conditions and healthcare systems affect disease results [2].

1.2 Electrocardiograms

An electrocardiogram (ECG) represents heart contractile activity as a signal which medical professionals obtain by placing electrodes on the chest or limbs. The ECG stands as the most widely recognized biomedical signal which medical professionals use daily to monitor heart function. The

relationship between heart function and the expression of the ECG trace has been described by Dubin [3]. The ECG signal tracks the temporal changes in electrical potential variations [4]. The ECG signal contains three main waveforms which are named P, QRS and T (Figure 1). Medical professionals use waveform duration and shape together with waveform distance to evaluate heart condition [5].

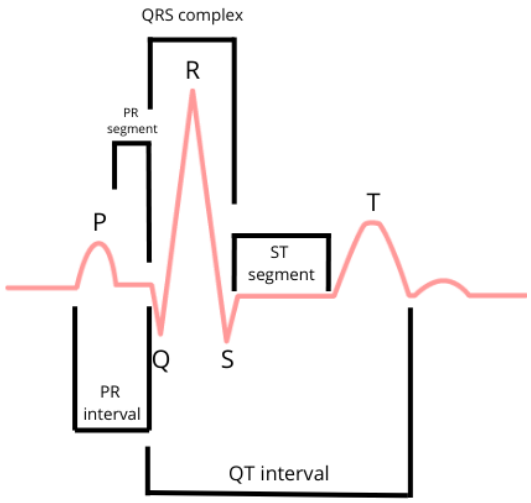


Figure 1: The plot illustrates the characteristic waveforms of a heartbeat cycle, including the P wave (atrial depolarization), the QRS complex (ventricular depolarization), and the T wave (ventricular repolarization)

The waveforms present various abnormalities which correspond to multiple cardiac conditions including arrhythmias and ischemia and myocardial infarction [6]. Ventricular conduction disturbance shows as a widened QRS complex while elevated or depressed ST segments indicate myocardial infarction or ischemia. The PR interval which measures the time between P waves and QRS complexes holds clinical significance because its prolongation can indicate AV block pathologies [6].

1.2.1 Ionic Basis of Cardiac Electrical Activity

Electrical activity in the heart originates because electrical impulses activate modifications between ions K^+, Na^+, Ca^{2+} and Cl^- outside and inside the myocardial cells. The positive charge accumulated on the cell membrane surface exists outside while the inner area maintains its negative charge because ions are dispersed unevenly [7].

1.2.2 Challenges in Traditional and Conventional ECG Analysis

Alarmingly extensive research has been directed toward automated ECG diagnostic systems because medical professionals need exact and speedy cardiac analysis. Traditional ECG interpretation methods face limitations because their depend on manual operation of cardiologists who may fall short in detecting subtle abnormalities from large data samples as highlighted by Acharya et al. [8]. The ability to automate some tasks through traditional signal processing methods advanced during recent years but these advances have failed to produce effective detection of rare and short-lived abnormalities [9].

1.3 Advancements in Healthcare Machine Learning: The Need for New Algorithmic Implementations

The extensive expansion of electronic health records containing complete digitized healthcare information has made machine learning (ML) possible for advancing healthcare innovation [10]. The analysis of large datasets has received additional support through efficient methods of analysis [11]. The development of new ML algorithms proceeds due to hardware advancements that include established computing platforms, including cloud computing, alongside high-performance systems and Graphics Processing Units and modern software systems [12].

1.4 Machine Learning Overview

The use of mathematical models with algorithms in machine learning allows tasks to attain improved performance through continuous enhancements [13]. The system uses training datasets while accepting information to produce decisions without requiring human coding. Machine learning contains two main task categories, which consist of supervised learning and unsupervised learning [13].

The goal of unsupervised learning is to detect patterns in input data that exist in datasets lacking pre-defined output labels [13]. Supervised learning functions using datasets containing inputs together with pre-defined outputs, and the semi-supervised learning subset works with incomplete output datasets. Supervised learning performs classification and regression tasks, yet unsupervised learning functions best when discovering new features and reducing data dimensions [13].

1.4.1 Limitations of Traditional Signal Processing and Machine Learning in ECG Analysis

Several prior research works examined different signal processing algorithms with machine learning solutions for improving ECG signal analytics. Discrete wavelet transform (DWT) and principal component analysis (PCA) represent two extraction methods that detect and classify arrhythmias by extracting vital information from ECG signals [14]. Most of these techniques depend on manually designed features although these features fail to fully capture the complex nonlinear behavioral patterns in ECG measurements. The diagnostic tool faces major obstacles because ECG signals show substantial variability between patients due to natural reasons and disease-related factors.

1.5 A Novel Approach for Accurate and Scalable ECG Diagnosis

The model that is going to be presented, utilizes solely the second ECG lead during training to detect cardiac conditions while conventional systems normally require all 12 leads for diagnosis. The simpler hardware requirements and portable usage capabilities, hypothetically will make the system operate well in wearable as well as remote care scenarios even when limited leads are accessible. The model should achieve high diagnostic accuracy despite its reduced input scope which addresses the traditional model failure during limited lead usages.

The system will provide a web-platform to retrieve ECG signals from digital pictures of displayed or printed ECG electrocardiogram graphs. Data entry is possible through this system which operates even without direct access to digital raw signals.

1.6 Objectives, Goals, and Research Questions

The study aim on solving the multi-label classification problem within ECG analysis. Multiple cardiovascular conditions tend to occur concurrently so the prediction model must demonstrate the capability to identify and handle several labels while understanding the related nature of different medical conditions. This research employs a classifier chain ensemble model as a solution to enhance the model's performance for handling complex multi-label classification activities in ECG diagnostic systems that face current limitations.

The main goal is to develop a web-based system that permits healthcare professionals working in distant or resource-limited areas to submit single-lead ECG images for instantaneous diagnostic output.

The main objective of this study is to train several models on a clinic ECG data’s and building a web-based system that integrates the developed machine learning model. The system provides

remote healthcare providers mainly from lower-resource and distant locations with an online platform to transmit single-lead ECG data which generates immediate diagnostic results. The system implements image-based data entry to provide automated ECG diagnosis while minimizing hardware dependencies which will increase access to ECG analysis diagnosis in situations lacking complete 12-lead ECG systems.

Multiple essential research questions guide the examine of these research objectives. How effectively can a hybrid machine learning model, combining Convolutional Neural Networks (CNN) and classifier chain ensemble methods, classify multi-label ECG data for cardiovascular abnormalities? Does the model demonstrate better diagnostic capabilities when it comes to diagnosing uncommon ECG disabilities including hypertrophy and conduction disturbances? What are the strengths and limitations of using machine learning models for ECG analysis?

Through these objectives and research questions, this study seeks to make a meaningful contribution to the field of automated cardiovascular diagnosis, providing a robust, scalable solution that can improve healthcare delivery worldwide.

CHAPTER 2- (BACKGROUND AND RELATED WORK)

The chapter introduces basic concepts and existing research that relates to this study to develop complete understanding of methodologies along with their identified gaps which this work will address. The initial part discusses ECG signal physiological components along with their clinical value in cardiac condition detection. A critique of traditional signal processing methods accompanies an evaluation of machine learning strategies as well as current automated ECG analysis approaches. This part of the paper points out shortcomings in current approaches to cover the way for the new method described in this research.

2.1.1 Physiological Basis of ECG Signals

ECG functions as a crucial tool for cardiovascular diagnosis because it detects essential information about heart electrical signals. Medical practitioners use the ECG tool, developed by Willem Einthoven in 1903, as an essential non-invasive device for heart rhythm evaluation, myocardial infarction detection, and cardiovascular condition monitoring [15]. A review of ECG signal physiology and their components, together with the abnormal findings they detect, utilizes research findings from the field.

2.1.2 Explanation of ECG Signals and Clinical Relevance

The ECG records bioelectrical signals that emerge from heart activities while depolarization and repolarization take place through each cardiac cycle. The sinoatrial (SA) node generates electrical

impulses that move along the conduction pathways, starting with the atrioventricular (AV) node, then the Bundle of His, and ending with Purkinje fibers, thus creating synchronized contractions between cardiac chambers [16].

The number of leads being recorded simultaneously determines whether an ECG falls under the single or multi-lead category. The most common systems utilize twelve ECG leads to display different cardiac action directions in three dimensions. The ECG leads have received specific names from I to III, aVR to aVF, and V1 to V6. The six limb leads of an ECG include I, II, III, aVR, aVL, and aVF, while the six precordial leads are V1 through V6 [15].

Medical personnel use the ECG as a crucial test without invasive procedures to detect heart rate together with rhythm and electrical conduction assessment. Studies show that deviations in these data points indicate potential cardiac issues, which include heart failure, ischemia, and atrial fibrillation [15]. The ECG functions as a critical diagnostic instrument within acute care environments to promptly recognize lethal heart conditions, which include myocardial infarction and arrhythmias.

2.1.3 Key Parameters of the ECG

The ECG waveform shows separate features that correspond to cardiac electrical periods and deliver vital information for diagnosis. Atrial depolarization creates the P wave which provides essential clinical information about atrial performance. Diagnostics through ECG recordings become possible when evaluation shows abnormal patterns or contrasts in the waveforms from normal characteristics because this can signal atrial enlargement together with arrhythmias. Ventricular depolarization appears as the essential QRS complex on electrocardiogram tracings. A heart rate of more than 120 beats per minute can lead to an appearance of a prolonged QRS width. According to Martis et al. [14], scientists must examine the QRS complex when they need to detect conduction abnormalities. The T wave shows how the ventricles complete the recharge process. Electrolyte

imbalances together with ischemia and ventricular strain normally appear through unique changes in T-wave shape including inversion or flattening. The ST segment demonstrates particular importance for diagnosing events of myocardial infarction because it marks ventricular repolarization. According to Zheng et al. [9], acute myocardial infarction detection requires examination of both ST-segment elevation and depression. A healthcare professional uses R-R interval measurements to study heart rate variability when assessing the duration of a complete cardiac cycle. The length of this period helps identify arrhythmias and autonomic dysfunction problems. ECG technology offers a unified perspective of cardiac electrical activity through its essential components so doctors can depend on it as their core diagnostic instrument for cardiovascular health needs.

2.2 Major Cardiovascular Disease Types

2.2.1 Hypertrophy

Under biomechanical stress various stimuli cause the heart to react through a process known as cardiac hypertrophy. Afflicted persons face an increased risk of sudden death or overt heart failure because of hypertrophy even though the condition can slowly adjust wall tension. Studies using both animal subjects and humans demonstrate that hypertrophy functions as an undesired biological process rather than normal homeostatic change from mechanical strain alterations [17].

2.2.2 Myocardial Infarction

A heart attack, also referred to as myocardial infarction (MI), results in heart muscle tissue death because of blood supply interruption. The clinical identification of this condition requires electrocardiographic changes, elevated biomarkers of myocardial necrosis, and imaging tests [18]. Worldwide statistics show that MI creates significant disability alongside mortality numbers, which often emerge as the initial clinical indicator of coronary artery disease [18].

2.2.3 ST/T Change

ST-T changes are recorded as electrocardiogram (ECG) segments that show myocardial ischemic or infarction rates through ST segment transformations along with T wave variations. The diagnostic indicators for diagnosing acute ischemic events depend on these heart signal alterations that involve both ST-segment elevation or depression and T wave inversion [19].

2.2.4 Conduction Disturbance

The electrical conduction system of the heart develops abnormalities that result in arrhythmias. An expert can detect these disturbances during an ECG test through which they identify conditions that include left or right bundle branch block which makes it harder to analyze myocardial infarction events [20].

2.2.5 Other abnormalities

Various abnormalities within ECG recordings, such as alterations in QRS complex readings, can lead medical professionals to misinterpret these patterns as indicative of myocardial infarction. These variations, including changes in QRS complex amplitude, duration, and distortions of its normal pattern, may also suggest conditions like ventricular hypertrophy or dilatation [21]. For instance, an increased QRS amplitude might point to left ventricular hypertrophy, often linked to hypertension or aortic stenosis, while a widened QRS complex could indicate bundle branch block or ventricular dilatation due to heart failure. Such misinterpretations can complicate diagnosis, as these abnormalities may mask or mimic the ECG signs of acute coronary syndromes, necessitating careful analysis by clinicians to differentiate between these conditions [15].

2.3 Automated approaches for ECG analysis

Automated approaches for ECG analysis have become increasingly important for the early detection and diagnosis of cardiovascular diseases. These methods leverage advanced machine learning and signal processing techniques to improve the accuracy and efficiency of ECG interpretation, offering a scalable solution to support clinicians in managing the growing burden of heart-related conditions.

Deep learning approaches, particularly convolutional neural networks (CNNs), have shown significant promise in ECG analysis. For instance, a novel Deep Multi-Scale Fusion neural network (DMSFNet) has been developed to enhance multi-class arrhythmia detection by capturing abnormal patterns and suppressing noise through multi-scale feature extraction. This method demonstrated superior performance on datasets like CPSC_2018 and PhysioNet/CinC_2017, achieving F1 scores of 82.8% and 84.1%, respectively [22]. Another study introduced a CNN combined with the Constant-Q Non-Stationary Gabor Transform (CQ-NSGT) to identify congestive heart failure and arrhythmia, reporting an accuracy of 98.9%, sensitivity of 98.8%, and specificity of 99.0% on the PhysioNet CHF dataset [21]. A deep neural network (DNN) was developed to classify 12 rhythm classes, achieving an average area under the ROC curve (AUC) of 0.97, surpassing the sensitivity of average cardiologists when tested on a dataset of ambulatory ECGs [23]. These advancements underscore the potential of automated systems to enhance diagnostic precision, often outperforming traditional methods and human experts in specific tasks, thus facilitating faster and more reliable cardiovascular care.

Recent research developed a deep learning detection method tailored for myocardial infarction (MI) and conduction disorders (CD) from the extensive PTB-XL ECG data base [25]. The study analyzed three significant issues that affect deep learning analysis of real-world ECG data

including its high computational requirements combined with limited model generality and poor interpretability of results. The system combined an optimized activation function which accelerated convergence with deep signal feature extraction through CNNs and then sent these data to an external support vector machine (SVM) for classification. The model displayed promising results as its overall accuracy reached 99.20% through the detection of MI and CD without depending on well-known datasets like MIT-BIH.

Table 1 shows diagnostic performance of multiple baseline and advanced ECG classification networks in terms of the PTB-XL test set. The performance metric values selected for comparison include Accuracy, Area Under the Curve (AUC), Recall, Precision, and F1-score. Each of the networks evidenced specific diagnostic abilities for ECG abnormalities, with sharp differences spotted when measured using different performance indicators.

Table 1. Diagnostic Performance Comparison of ECG Classification Models on the PTB-XL Test Dataset

Model	**Accuracy**	**AUC**	**Recall**	**Precision**	**F1-score**
ResNet-Wang[24]	0.877	0.749	*0.795*	0.712	0.751
ECG-DNN [25]	0.884	0.924	0.684	*0.793*	0.734
LightX3ECG[26]	0.884	0.920	0.681	*0.795*	0.734
2D-ECGNet[27]	*0.892*	0.929	0.790	0.752	*0.770*
CBMV-CNN [28]	0.880	*0.935*	*0.796*	0.725	0.759
CAMV-RNN [28]	0.889	0.932	0.783	0.748	0.765
STFAC-ECGNet [28]	*0.894*	*0.933*	0.756	0.778	*0.767*

2.3.1 Limitations of Automated Approaches in ECG Analysis

The ability of traditional machine learning systems in ECG classification remains poor right now particularly compared to deep learning. The implementation of handcrafted features in ML models produces limitations in detecting complete ECG waveform complexity [15]. The inability to identify weak patterns associated with myocardial infarction or conduction disorders causes these

systems to fail their diagnostic duties. The performance of numerous ML-based systems becomes reasonable only when they have access to all 12 ECG leads [24]. The limited number of available ECG leads in wearable devices and remote care excludes their practical application because these models function most effectively with multiple leads. The diagnostic reliability suffers because models which receive full 12-lead data training fail to retain their functionality when only limited leads are utilized. The current difficulties demonstrate the requirement for new advanced and lead-unspecific solutions within automated ECG diagnostic systems.

The use of various ECG devices together with different sampling rates and leads creates uncertainty in automated systems because it impacts the signal quality [26]. The lack of explainability features in numerous models prevents doctors from comprehending the purposes behind specific predictions [27]. The lack of clear reasoning behind results makes healthcare providers withhold confidence when interacting with these results. Even though ongoing technical advancements occur the present ECG devices show several limitations within their manual and automatic operational areas.

CHAPTER 3 - (MATERIALS AND METHODS)

3.1 Dataset

The PTB-XL diagnostic ECG dataset which represents a publicly available dataset was selected for the train and test set. A total of 21,799 10-second clinical 12-lead ECG recordings exists in the PTB-XL diagnostic ECG dataset that includes measurements from 18,869 patients spanning from 0 to 95 years of age (median: 62 and interquartile range: 22) [29]. The dataset consisted of 52% male patients while the other 48% were female patients. The main worth of this dataset comes from its wide coverage of cardiovascular coexisting conditions alongside a sizeable number of health control cases.

The diagnostic labels were restricted to aggregated diagnostic superclasses to improve analytical efficiency since individual diagnostic statements surpass the record count when multiple diagnoses occur in a single ECG recording. Table 1 shows the distribution summary of ECG records according to their diagnostic superclasses.

Table 1. Distribution of ECG Records by Superclass and Condition Description.

Number of records	Superclass	Description
9514	NORM	Normal ECG
5469	MI	Myocardial Infarction
5235	STTC	ST/T Change
4898	CD	Conduction Disturbance
2649	HYP	Hypertrophy

Before preprocessing started the PTB-XL database was reorganized to ensure data consistency while design simplification for model training purposes. Multiple diagnostic statements appear under each ECG record in the original dataset format which provides multi-label annotations. But this arrangement turned the problem into a complex multiclass-multioutput system. The management of this issue involved grouping diagnostic labels into larger superclasses which included medically related conditions. Every superclass had associations with different diagnostic subclasses. Three subclasses named AMI, IMI and ILMI exist within the MI superclass and the CD group includes the conduction disorders LAFB, IRBBB and 1AVB. The mapped structure helped the model to perform a focused learning process and simultaneously reduced label confusion and the presence of unnecessary details. The resultant dataset achieved better balance which made it ready for training a multi-class classifier. The diagnostic subclasses in Table 2 receive assignment to superclasses through the specified mapping methodology.

Multiple diagnostic labels remained ambiguous regarding their assignment to existing superclasses when the subclass aggregation procedure took place. The resulting ABNORM category collected undefined ECG conditions. The ABNORM category functioned as a catch-all for unclassified but abnormal ECG conditions whose subclasses joined together in this specific

classification group. The method stopped valuable diagnostic information from being removed so all records stayed active within the training database. The structured subclass system enabled the model to focus its learning activities and eliminated both vague labels and superfluous information. The improved dataset balance made the dataset appropriate for training multi-class classification models. This mapping procedure assigned diagnostic subclasses in Table 2 to their appropriate superclasses.

Table 2. Mapping of Diagnostic Subclasses to Aggregated Superclasses in the PTB-XL Dataset.

Superclass	Diagnostic Subclasses
NORM	NORM
MI	IMI, ASMI, ILMI, AMI, LMI, ALMI, IPLMI, IPMI, PMI, INJLA
HYP	LVH, LAO/LAE, RVH, SEHYP, RAO/RAE, VCLVH
CD	LAFB, IRBBB, 1AVB, IVCD, CRBBB, CLBBB, LPFB, ILBBB, WPW, 3AVB, 2AVB
STTC	NST_, DIG, LNGQT, ISC_, ISCAL, ISCAS, ISCLA, ISCAN, ISCIL, INJAS, INJAL, INJIN, INJIL, EL, ANEUR, TAB_, NDT, STD_, STE_, INVT, LOWT, ISCIN
ABNORM	LVOLT, HVOLT, ABQRS, AFIB, PSVT, SBRAD, LPR, STACH, PRC(S), QWAVE, SVTAC, PAC, AFLT, SR, SVARR, NT_, SARRH, PVC, BIGU, TRIGU

3.1.1 The workflow of the proposed method:

The method begins with signal preprocessing through a filtering operation that optimizes signal quality. Both baseline drift removal and powerline interference suppression occur through filtering signals with a 0.5 Hz high-pass Butterworth filter followed by a 50 Hz notch filter. The signal processing uses the Pan-Tompkins algorithm after removing noise components to enhance QRS complex detection. The algorithm begins by filtering signals then applies differential computation and square operation while conducting moving window integration to detect and recognize the QRS complex effectively.

After preprocessing the ECG signal is processed by Discrete Wavelet Transform (DWT) to detect the essential fiducial points P-wave, QRS complex, and T-wave. The necessary temporal attributes for analysis (PR interval, QT interval, RR interval, ST segment and QRS duration and P-wave and T-wave durations) are derived from these fiducial points. The conversion to time (milliseconds) happens through application of the sampling frequency to the values.

A computer program extracts peak amplitudes of P, Q, R, S, T components in ECG data because these measurements aid in detecting cardiac conduction problems while identifying additional abnormal conditions. The mean heart rate calculation comes from RR intervals and the cardiac phase analysis differentiates atrial from ventricular activity while phase durations normalize against RR interval differences to handle heart rate changes.

The structured data frame contains time-domain metrics along with amplitude-based metrics as well as phase-related metrics that have been extracted from the data. The process for data consistency includes value cleaning which allows statistical calculations of mean and standard deviation for each feature.

The Z-score method enables outlier detection through standardization of features to detect

important deviations from the mean values before removal. Extreme values will be removed by this step because they could damage both the accuracy of statistical tests and degrade the performance of the model.The extracted features undergo evaluation through 11 machine learning models which belong to multiple families for classification purposes. 5-fold cross-validation serves as the assessment method to confirm generalizable results.

A Z-score standardization method is the final step which makes all features equally comparable to each other when applied to machine learning models.

3.2 Preprocessing

The signal quality enhancement process begins the preprocessing step. An initial filtering method helps eliminate unwanted artifacts before optimizing ECG signal clarity. Users apply high-pass filtering with Butterworth design at 0.5 Hz and 5th order cutoff to stop baseline drift from occurring. Powerline interference suppression occurs through the application of a 50 Hz notch filter to protect the signal from high-frequency electrical noise degradation. The Pan-Tompkins algorithm [30] operates after preliminary noise elimination to boost QRS complex detection of the signal. Accurate R-peak classifications depend on this method because it emphasizes how the fast cardiac depolarization process distributes its frequencies. Pan-Tompkins starts by filtering out background noise except for QRS complex and eliminating P- and T-waves. This process differentiates the signal to show the rapid slope changes in QRS complexes. A squaring function enhances high-frequency wave components so the QRS complex becomes more noticeable but it simultaneously reduces the influence from other waveform elements. The detection of QRS complexes depends on smoothing the squared signal with an integration function that uses moving windows for optimization. Dynamic thresholding method uses signal-adaptive analysis of the

processed input to detect R-peaks precisely in various ECG conditions. Figure 2 visualizes how Pan–Tompkin's algorithm prepares the signal for feature extraction by showing the progression of sequential stages that refines the signal.

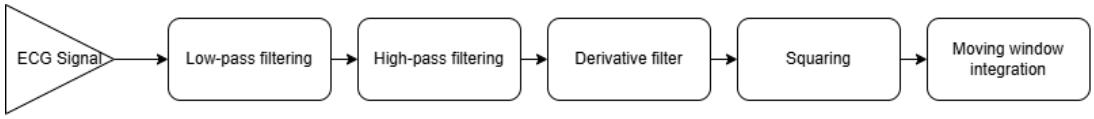


Figure 2. Block diagram of the pre-processing phase of the Pan–Tompkins algorithm.

3.3 ECG Feature Extraction: Peak Identification and Interval Analysis

The analysis of ECG waveform detects vital markers known as fiducial points that include P-wave, QRS complex and T-wave onset and peak and offset. A multiscale analysis technique called Discrete Wavelet Transform (DWT) enables precise location precision of features. The following table describes fiducial points along with their physiological meaning (Table 3).

Table 3. Overview of key ECG wave components, their corresponding fiducial points, and the physiological events they represent during the cardiac cycle.

Wave Component	Fiducial Point	Physiological Significance
P-wave	P_peak	Atrial depolarization peak
P-wave	P_onset	Start of atrial depolarization
P-wave	P_offset	End of atrial depolarization
Q-wave	Q_peak	Initial downward deflection of QRS complex
S-wave	S_peak	Terminal downward deflection of QRS complex

T-wave	T_peak	Ventricular repolarization peak
T-wave	T_onset	Start of ventricular repolarization
T-wave	T_offset	End of ventricular repolarization

3.3.1 Temporal Feature Computation

To extract key temporal features from the ECG signal, the identified peak indices are converted into time values based on the sampling frequency ($\boldsymbol{f_s}$). Since the peak indices are initially measured in sample points, they are converted into milliseconds **($\boldsymbol{ms}$)** using the following equation:

$$T = \frac{P}{f_s} \times 1000$$

Where T represents the time in milliseconds, P denotes the peak index (i.e., the sample number at which the peak occurs), and f_s is the sampling frequency in Hz (samples per second). The factor of 1000 ensures conversion from seconds to milliseconds. Using this approach, key ECG time intervals and segments are calculated.

The **PR interval**, which represents the time between the onset of atrial depolarization and the R-peak of the QRS complex, is computed as:

$$PR_{interval} = (R_{peak} - P_{onset}) \times \frac{1000}{f_s}$$

The **QT interval**, defined as the duration between the onset of the Q-wave and the peak of the T-wave, is given by:

$$QT_{interval} = (T_{peak} - Q_{peak}) \times \frac{1000}{f_s}$$

The **RR interval**, which denotes the time interval between successive R-peaks and is essential for heart rate variability (HRV) analysis, is calculated as:

$$RR_{interval} = (R_{peak,i+1} - R_{peak,i}) \times \frac{1000}{f_s}$$

The **ST segment**, which indicates ventricular repolarization and is measured as the time difference between the S-wave and the onset of the T-wave, is determined using:

$$ST_{segment} = (T_{peak} - S_{peak}) \times \frac{1000}{f_s}$$

The **PR segment**, representing the time between the end of the P-wave and the onset of the Q-wave and reflecting atrioventricular (AV) conduction, is computed as:

$$PR_{segment} = (Q_{peak} - P_{offset}) \times \frac{1000}{f_s}$$

The **QRS duration**, which corresponds to the total duration of the QRS complex and is essential for assessing conduction abnormalities, is calculated as:

$$QRS_{duration} = (S_{peak} - Q_{peak}) \times \frac{1000}{f_s}$$

The **P-wave duration** and **T-wave duration**, representing the time from the onset to the offset of each respective wave, are computed as:

$$P_{duration} = \left(P_{offset} - P_{onset}\right) \times \frac{1000}{f_s}, \quad T_{duration} = \left(T_{offset} - T_{onset}\right) \times \frac{1000}{f_s}$$

3.3.2 Amplitude-Based Feature Extraction

Medical professionals can directly interpret peak amplitudes of P, Q, R, S, T components from the processed ECG signal. The detection of conduction problems and ischemic pathology as well as other pathological conditions depends on accurate measurement of these amplitudes.

3.3.3 Heart Rate and Cardiac Phase Analysis

The mean heart rate is computed using the formula:

$$HR_{mean} = \frac{60000}{\overline{RR_{interval}}}$$

where:

HR_{mean} is the mean heart rate in beats per minute (bpm),

$\overline{RR_{interval}}$ is the average RR interval in milliseconds,

The factor 60000 converts milliseconds to minutes since heart rate is measured in beats per minute.

3.3.4 Cardiac phase extraction.

Cardiac phase analysis is performed to differentiate atrial and ventricular activity. Using delineated ECG waveforms, the atrial and ventricular phases are extracted.

Cardiac phase extraction quantifies atrial and ventricular activity by assigning a phase value to each point in the ECG signal. The atrial phase duration (D_{atrial}), which corresponds to the period of atrial depolarization and repolarization, is measured as the time interval from the onset of the P-wave to the offset of the P-wave, given by:

$$D_{atrial} = P_{offset} - P_{onset}$$

The ventricular phase duration ($D_{ventricular}$), representing the total duration of ventricular

depolarization and repolarization, is computed as the time interval between the onset of the Q-wave and the offset of the T-wave:

$$D_{ventricular} = T_{offset} - Q_{onset}$$

Cardiac phase durations, D_{atrial} and $D_{ventricular}$ are measured in absolute time (milliseconds). Since higher heart rates correspond to shorter RR intervals, phase durations tend to decrease as heart rate increases. Conversely, lower heart rates lead to longer phase durations. This inverse relationship introduces variability that is unrelated to intrinsic cardiac conduction properties, making it difficult to determine whether observed differences in phase durations are due to actual physiological changes or simply variations in heart rate.

To mitigate this heart rate dependency, normalization is applied by expressing phase durations relative to the total cardiac cycle duration.

This is achieved by dividing each phase duration by the corresponding RR interval, ensuring that phase durations are analyzed proportionally to the full heartbeat. The normalized atrial and ventricular phase durations are computed as:

$$D_{atrial,\ norm} = \frac{D_{atrial}}{RR_{interval}}, \qquad D_{ventricular,\ norm} = \frac{D_{ventricular}}{RR_{interval}},$$

This normalization approach allows for more accurate comparisons across individuals and conditions by reducing the influence of heart rate variability on phase duration measurements.

3.3.5 Feature Compilation and Output

A structured data frame contains all extracted features which include time-domain measures together with amplitude-based and phase-related metrics. Protection of data consistency requires that invalid values such as NaNs get eliminated while indices get converted to integer

format. The extracted peaks alongside all intervals pass through an adjustment system which sets them to match the shortest possible extraction length for uniform results.

The analysis produces 36 derived features by applying standard deviation measurements on the extracted values alongside mean calculations. There are two main measurements used for each feature in this approach. Firstly, the mean calculates average measurements (e.g. time or amplitude) and secondly, standard deviation quantifies the measurement variability.

3.3.6 Outlier Detection and Removal

To protect data consistency and reduce the effects of unusual values in model evaluation and training processes outlier detection with removal mechanism was implemented on extracted ECG features. A data point deviation measurement called Z-score enabled the identification of outliers from the mean value based on standard deviations.

The set of numerical features from ECG waveform delineation received the outlined procedure. The standardization procedure was applied to all features using the Z-score method. The absolute Z-score calculation took place sample by sample while considering all the features present in each row. The statistical analysis defined 2.5 as the threshold value which detected outliers when any sample had at least one feature with an absolute Z-score greater than that value.

These unusual observations were removed from the dataset because they stemmed from measurement artifacts or noise and possible rare physiological conditions. The removal of features that exceed the threshold value of 2.5 helps to protect data distribution integrity while decreasing bias and overfitting risks throughout the model learning process. The cleaned dataset contained only physiological samples after removal allowing statistical analyses to become more reliable and enhancing model generalization capabilities.

3.3.7 Handling of Missing Values

The data processing involved a sequential approach for imputing feature matrix gaps to support uninterrupted model training.

The initial step of analysis removed dataset columns containing only NaN entries from the entire matrix. The irrelevant columns which provide zero value for classification actually reduce the performance efficiency of later modeling procedures.

The mean imputation process was applied to fill gaps in all features except those removed previously. Each missing value became replaced by the mean value of the feature using Simple Imputer with "mean" strategy over all available samples. The method maintains all present features' original distribution patterns yet minimizes errors caused by missing data points.

3.3.8 Feature Scaling

To normalize the range of feature values and ensure comparability across different measurement units, *Min-Max scaling* technique was applied to the feature matrix. This preprocessing step rescales each feature to a fixed range of [0, 1], preserving the relative relationships among values while eliminating disparities in scale.

A *MinMaxScaler* from the `scikit-learn` library was employed to perform this transformation. The scaler computes the minimum and maximum values for each feature and applies the transformation:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

The scaling was applied after imputation and outlier removal, ensuring that the normalized features accurately reflect the cleaned and complete dataset.

3.3.9 Multi-Label Resampling via REMEDIAL and MLeNN

The two-phase resampling strategy was utilized as a solution to handle the typical problems of imbalanced classes and label repeated occurrences found in multi-label ECG classification. The framework unites an advanced version of REMEDIAL along with MLeNN after modifying them for minority-class balance management along with signal diversity preservation.

a) REMEDIAL: Oversampling of Minority Classes with Label Decoupling

The REMEDIAL algorithm in the first phase utilized REsampling MultilabEl datasets by Decoupling highly ImbAlanced Labels approach from Charte et al. [31] to decrease the frequency of minority and majority class labels within individual samples. Two measurement scales served as guiding factors for this process:

- IRLbl (Imbalance Ratio per Label): quantifies label imbalance by comparing the frequency of each label to the most frequent one.
- SCUMBLE (Score of ConcUrrence among iMBalanced LabEls): measures how often minority labels appear alongside majority labels, which can obscure minority-class learning [31].

High SCUMBLE scores evolved into distinct instances that included either minority classes exclusively or majority classes independently similar to label separation. During resampling the minority class samples received additional duplications to achieve balance representation against underrepresentation. The regulation parameters received adjustments through scumble_multiplier set to 0.9 and irlbl_multiplier set to 0.9 to implement a stronger overlapping label context separation method.

b) MLeNN: Undersampling of Noisy or Ambiguous Samples

The experiment utilized MLeNN algorithm variation from Charte et al. [32] for noise-filtering after decoupling and oversampling steps. The algorithm conducts undersampling on instances through a label agreement evaluation mechanism that involves each sample with its k-nearest neighbors. The samples with significant disagreement levels get classified as noisy or mislabeled and undergo removal.

The implementation applied Jaccard similarity for label set agreement evaluation through a 0.9 disagreement threshold alongside k = 3 neighbors. The method received an improvement which included minority-class preservation to keep samples containing infrequent labels (IRLbl-score identified) after detecting disagreement occurred.

c) Outcome

The combined REMEDIAL oversampling approach for minority-class cases alongside MLeNN underrepresentation of ambiguous major-class patterns generated a training dataset which was more balanced and knowledgeable about cardiac signal variations. The classifier succeeded in identifying subtle ECG signal variations of rare cardiac conditions because of this approach.

3.4 Machine Learning Models

3.4.1 Machine Learning Models for ECG Feature Classification

Several commonly-utilized machine learning models formed the basis of traditional classification algorithms to evaluate the discriminatory power of extracted ECG features. A focused assessment of base-level classifiers was possible after selecting the models from alternative categories of ensemble and regression approaches.

Linear Discriminant Analysis (LDA) belongs to category one linear classifiers although physicists frequently use it because its efficient operation and simple interpretation works well for biomedical applications.

The conditional independence assumption of features makes Naïve Bayes probabilistic models work by using Bayes' theorem to find class probabilities. These classifiers function quickly while staying resistant to signal deviations.

Support Vector Machines (SVM) belong to the kernel-based methods which establish optimal separating hyperplanes through feature space transformations to deal efficiently with complex and overlapping class boundaries.

Classification and Regression Trees (CART) perform recursive partitioning of features to establish hierarchical rule-based classifiers which work as tree-based models. The models provide explanations while keeping the ability to detect non-linear class boundaries.

Instance-based classifiers operate through the KNN method that divides new instances by identifying the neighboring features with assigned labels in the feature space. The KNN model demonstrates specific strength in recognizing classification patterns in datasets that show significant interpretation at the local level.

The scikit-learn library (version 1.6.1) provided default hyperparameter settings to

execute all models with standardized evaluations that eliminated possible biases which could have come from manual model adjustments. The uniform framework allows performance comparison between different algorithms which utilize extracted ECG features for multi-class categorization.

3.4.2 Ensemble Models

Multiple ensemble learning methods were used in this research to enhance both accuracy and robustness levels in ECG classification. Multiple individual predictive models assemble into an enhanced reliable predictive system with ensemble methods.

XGBoost (Extreme Gradient Boosting) served as the main technique alongside being an efficient and scalable implementation of gradient boosting. XGBoost operates quickly and precisely by implementing weak decision trees to generate one powerful learning system. Through its optimization process the algorithm manages between regularization and loss function performance to stop the development of overfitting. The ability of XGBoost to handle structured/tabular data enables high predictive capabilities for classification tasks because it works effectively with ECG features.

The model application of Gradient Boosting depended on creating weak learners sequentially to construct an ensemble. The latest model version attempts to resolve previous model issues through optimized data pattern detection. XGBoost provides excellent bias reduction yet needs professional parameter adjustment to prevent model overfitting.

The ensemble method referred to as Random Forest implements bagging (Bootstrap Aggregating) to derive its predictions. Random Forest constructs several decision trees which learn from distinct pieces of the available data. The model generates predictions through collective analysis of numerous trees to enhance its stability and minimize error variance. The method offers

a robust and simple approach which enables efficient processing of large datasets containing extensive features thus making it highly effective for ECG signal classification.

The system incorporated Light Gradient Boosting Machine (LightGBM) as its speed and memory efficiency surpass those of gradient boosting. The efficient handling of big datasets by LightGBM occurs through its use of histogram-based algorithms which simultaneously quicken training speed and minimize memory requirements. LightGBM performs exceptionally well with categorical data and provides very fast data processing which makes it suitable for immediate decision-making tasks like real-time ECG diagnosis.

The Voting Classifier technique combines XGBoost and Random Forest with other models through a majority vote algorithm for classification tasks yet using averages predictions for regression tasks. The technique allows different classifiers to work together to minimize their individual biases while increasing the prediction accuracy.

The advanced ensemble method Stacking applies multiple base learners (XGBoost, Random Forest and LightGBM) to produce base model outputs that later become training data for creating a meta-model. The meta-model develops the capacity to combine base model predictions effectively for superior outcome accuracy. Fusion processing through Stacking creates superior performance than independent base models since it both retrieves hidden data segments and rectifies original model errors.

These ensemble techniques applied to ECG classification work collaboratively with PTB-XL data to improve model performance regarding dataset complexity and data consistency and imbalance issues. Multiple machine learning models together create better generalization and more dependable predictions which solve the issues identified in single-model applications.

3.5 Classifier Chain and Final Model Selection

The best model was selected based on maximal accuracy along with universal performance across the data following training simple and ensemble models. A Classifier Chain model was trained using 10 chains to deal with the multi-label requirements of ECG classification purposes. Multiple binary classifiers form a chain in sequence-based classifier chain models which utilize predictions from preceding labels as extra features for prediction purposes. The model benefit from label relationship detection because of its design which boosts performance when handling tasks with multiple label assignments. The model received additional stability through ensemble techniques when taking the average of predictions from its ten classifier chains to achieve better overall outcome performance. An independent model implemented the One-vs-Rest approach for training purposes as a separate evaluation method. The OvR strategy enables the model to analyze each category separately thus making it suitable when dealing with unbalanced data or non-discriminatory class boundaries. Multiple strategies merged within the final model delivered enhanced diagnostic accuracy which also proved effective at dealing with the ECG diagnosis task's multi-label characteristics.

3.5.1 Cross-Validation for Model Evaluation

The machine learning models went through training using the 10-fold cross-validation method. This technique organizes the data into 10 identical partitions called folds so that each subset works as the testing group once in the training process. The training activity occurs within the training set and follows with testing operations executed against the related test set. The procedure is implemented ten times to use every subset once for both training and testing purposes.

3.6 Hybrid Deep Learning and Machine Learning Approach

The ECG signals underwent high-level feature extraction using a deep learning model after which traditional machine learning models received this information. An assessment was conducted to select the optimal traditional machine learning model while relying on extracted features. A comparison of results between the classifier chain method and deep learning with combined machine learning models allowed to identify best methods for handling multi-label ECG classification tasks. The evaluation allowed to measure how well learned representations performed compared to manually designed features and features produced by models.

The 1D Convolutional Neural Network (CNN) served as a framework for extracting features from ECG data which involved classification tasks. A cascading set of three convolutional layers obtains features through filters of increasing dimension from 64 to 128 to 256 while using a kernel size of 3 with ReLU activation. Max-pooling operations occur in succession after each convolutional block to lower the size of extracted feature dimensions. The flattened output goes through two dense layers with 128 and 64 neurons that apply ReLU activation. A dropout layer with a rate of 0.5 comes after the initial dense layer to prevent overfitting.

Multi-class classification through sigmoid activation enables the last layer to predict one neuron for each class while training occurs with the binary cross-entropy loss. The training process takes place with Adam optimizer and 0.001 learning rate and 120 epochs and 32 batches adopting 20% training data as validation material. Both accuracy and loss indicators serve to track training performance. The model produced feature vectors to help classification techniques during completion of training while using the best-performing model trained from prior stages for final classification.

3.7 Web Platform Development for ECG Data Input

The web platform delivered a solution for simple uploading and analytical processing of ECG data. Through its User Interface (UI) medical staff can interact with the model. Users can use the system to upload ECG image recordings which include single-lead ECG signals taken as image files longer than 10 seconds. Interface was build using React for its front-end element. FastAPI working with Uvicorn processes requests for model interaction through its back-end infrastructure.

The platform allows ECG data input through image files of ECG signals that can be JPEG PNG. Users utilize the platform interface to directly upload single-lead ECG image recordings through this application. The platform starts a reading process of the uploaded ECG image through OpenCV library functionality. The imaging software retrieves only the red channel from the input data because traces in the ECG routinely have darker tone values in that spectrum. A binary mask gets created through red channel thresholding which separates the waveform from surrounding background. The binary mask reveals the ECG signal although it might include anomalous elements such as printed grid lines or artifacts. The vertical scanning process enables detection of pixel-rich columns which subsequently get excluded from image processing.

The system performs connected component analysis to discover all individual areas in the image. From these, only the largest component—typically the ECG trace—is retained. A morphological operation applies to the filtered mask to create a skeletonized trace as a single-pixel-wide line that improves extraction capabilities.

Supplemented by the skeletonized trace the program restores the signal in vertical progressions. Every image vertical position generates a point between 0 to 1 that gets saved into a one-dimensional signal array. Image vertical measurement corresponds to voltage which makes the signal need inversion and normalization to 0–1 values before model input. After extracting the

raw ECG signal, preprocessing stage, according to previous sections, starts to happen.

Chapter 4 – (RESULTS)

4.1 Five-Class Classification Performance

4.1.1 Conventional models

Five popular machine learning algorithms were employed to assess the discrimination potential of extracted ECG characteristics: Linear Discriminant Analysis (LDA), Naïve Bayes and Support Vector Machine (SVM) and Classification and Regression Tree (CART) as well as k-Nearest Neighbors (KNN). The performance metrics included accuracy together with F1-weighted score along with Hamming loss and area under the ROC curve (ROC AUC) assessment for each model. The complete results from Figure 3 include error bars showing standard deviation which visualize results across every metric.

The dataset demonstrated strong classification performance at a level of 0.68–0.69 for both KNN and SVM regarding accuracy attainment. The performance metric of accuracy measured LDA at 0.65 and CART along with Naïve Bayes exhibited lower values of approximately 0.65 and 0.53 respectively. Naive Bayes suffers reduced accuracy because its feature independence assumption does not align with the characteristics of ECG signal data.

All models demonstrated consistency in their measured F1-weighted score because this method addresses class imbalance and ensures balanced precision and recall results. The LDA model achieved maximum performance reaching 0.69 while SVM scored slightly lower at 0.66 and CART with KNN occupied the range from 0.66 to 0.68. The obtained results indicate that performance discrepancies did not impact how models allocate false negative versus false positive

occurrences.

The performance of both SVM and LDA regarding Hamming loss evaluation was optimal with readings at 0.09–0.10 although Naïve Bayes showed higher misclassification rates with a loss of 0.18. Naïve Bayes achieved the highest rates of multi-label misclassifications due to its relatively high Hamming loss of approximately 0.18.

The performance level of Naïve Bayes surpassed other models in the ROC AUC evaluation as it reached close to 0.77 points. The scores of LDA, CART and SVM matched in the same range of 0.70 to 0.72. The performance of KNN reached an AUC level of close to 0.68 which indicated limited capability to position positive classes higher than negative ones in multi-label scenarios.

The results show that KNN and SVM outperformed LDA during accuracy testing and Hamming loss evaluation although LDA demonstrated stable performance in all metrics. Naïve Bayes delivered better ranking performance than other models according to its ROC AUC although its basic classification precision was limited. The effectiveness of different algorithms is shown along with their limitations through these results while showing the necessity to use multiple evaluation metrics for assessing multi-label classification performance in ECG diagnostic tasks.

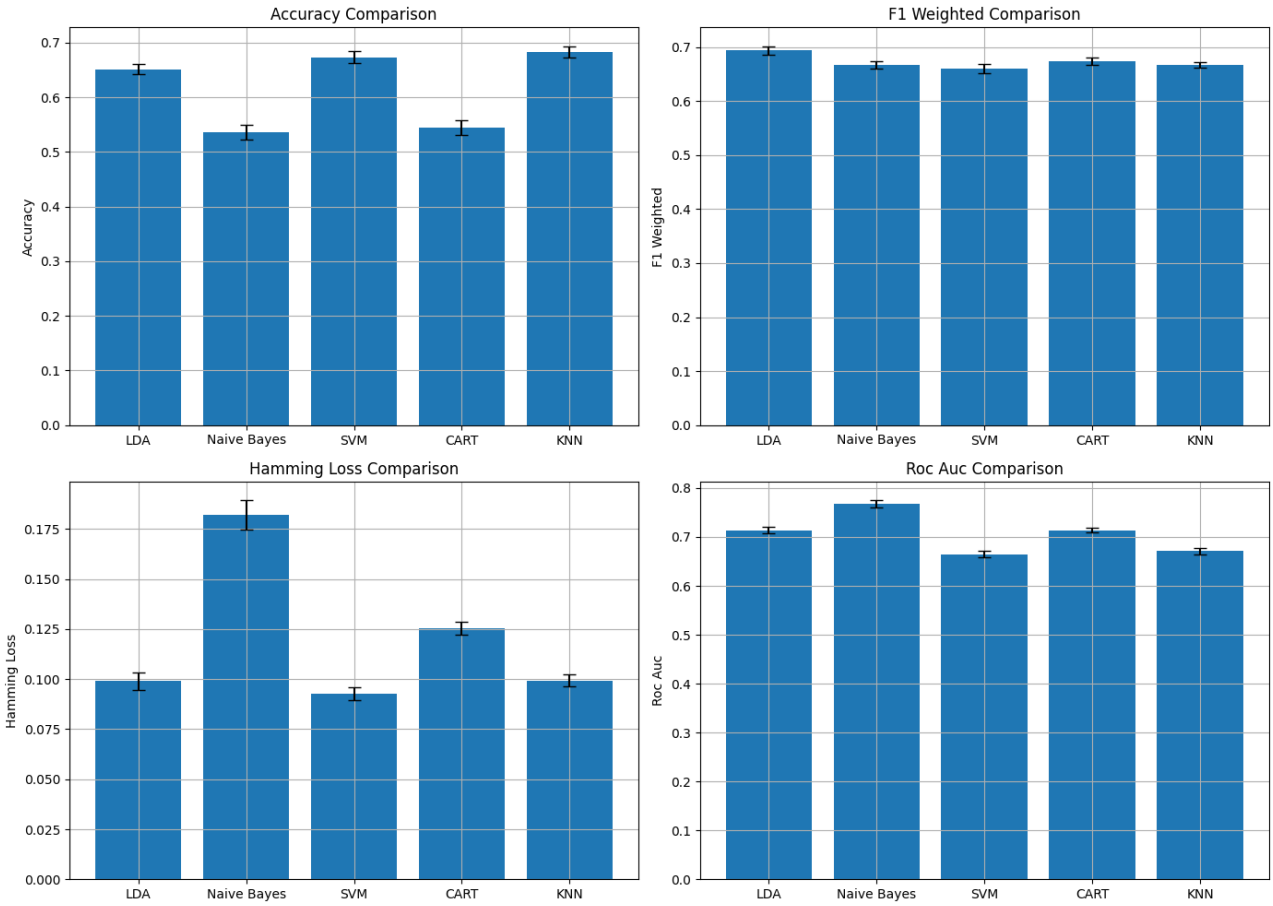


Figure 3. Comparative Performance of Baseline Classifiers Across Four Evaluation Metrics

4.1.2 Ensemble Classifiers Performance Evaluation

The evaluation of ensemble learning-based classifiers on extracted ECG features included Random Forest, Gradient Boosting, XGBoost, LightGBM, Voting and Stacking through performance testing. The described models were evaluated through four measures including accuracy and F1-weighted score and Hamming loss and area under ROC curve (ROC AUC). Figure 4 shows the testing results which include standard deviations from cross-validation folds.

Database accuracy reached similarly high levels among ensemble classifiers by staying between 0.66 and 0.71. The accuracy range of Random Forest together with XGBoost and LightGBM and Stacking reached 0.70–0.71 because these methods successfully utilized ECG

signal feature representations. Voting as well as Gradient Boosting achieved accuracy slightly lower than other ensemble classifiers reaching 0.66.

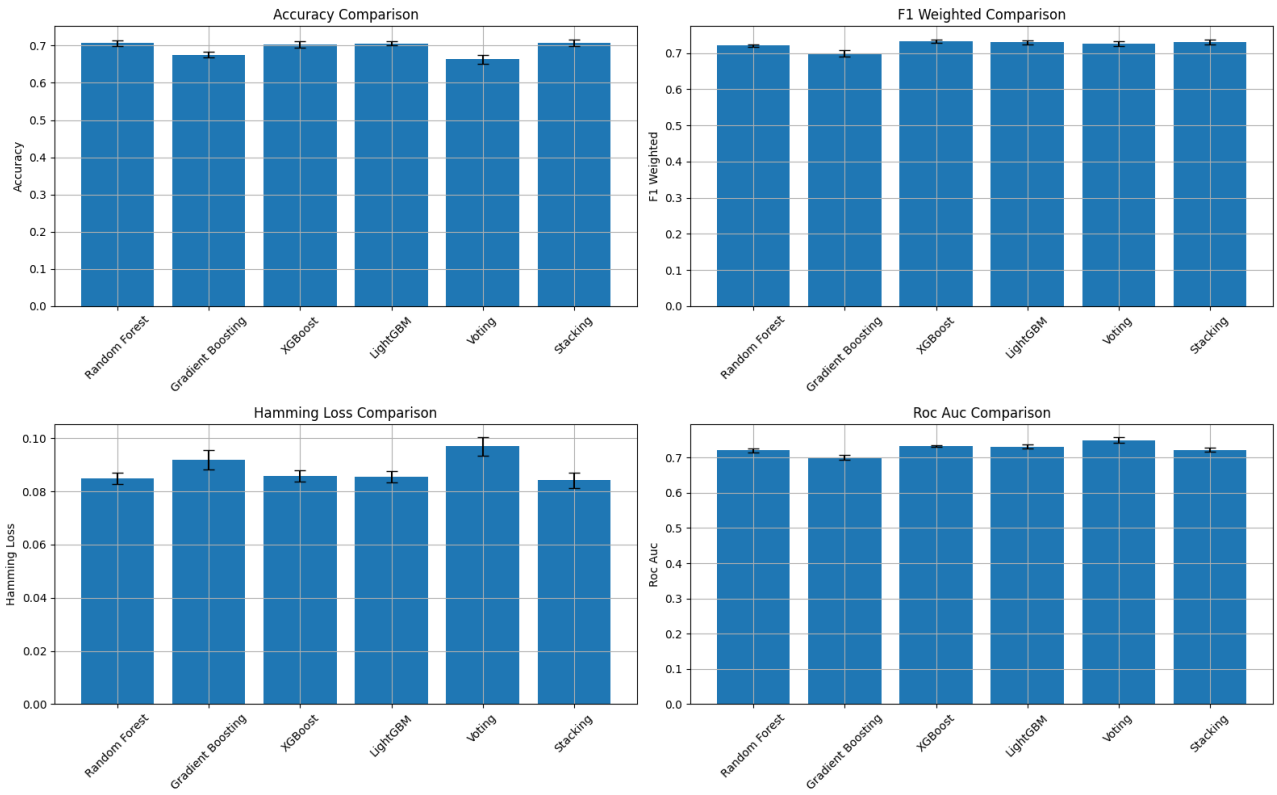


Figure 4. Comparative Performance of Ensemble Classifiers Across Four Evaluation Metrics

The F1-weighted score demonstrated regularly high values among all models because it both corrects class imbalance and strikes a balance between precision and recall. XGBoost and Stacking yielded the optimal results with F1 scores exceeding 0.73 while being matched by LightGBM and Voting at slightly lower scores. The predictive accuracy results demonstrate that these models provided efficient output while upholding equilibrium when making class predictions.

The Hamming loss evaluation demonstrated that both Stacking and LightGBM reached

the best results with losses at approximately 0.085 in the multi-label classification task. XGBoost and Random Forest showed similar performance numbers after XGBoost and LightGBM in second place. The slight increase in Hamming loss indicates Voting and Gradient Boosting produced a slightly higher frequency of wrong labeling outputs for their predictions.

Voting achieved the best ROC AUC score of 0.75 along with XGBoost and LightGBM which tied at this level. The models provided strong performance based on AUC metrics where Random Forest along with Stacking trailed XGBoost by a small margin and Gradient Boosting maintained an acceptable performance level after Random Forest.

All assessment metrics demonstrate steady high performance from ensemble classifiers as XGBoost together with LightGBM along with Stacking achieve consistent results throughout the analyses. Despite its basic structure the Voting classifier showed solid ROC AUC results which confirms its efficacy to merge diverse models together. Research data demonstrates that ensemble techniques offer remarkable potential for ECG-based prediction tasks while handling class imbalance effectively and building predictive resistance.

4.1.3 Performance Comparison and Statistical Analysis

The p-value obtained from an unpaired two-tailed t-test analysis of top model performance equaled 0.75 and exceeded the 0.05 significance threshold. The 10-fold cross-validation results indicate that LightGBM achieved the best performance but alternative models such as Bagging and Gradient Boosting demonstrated comparable results since their performance difference was not significant. A statistical analysis shows that the performance difference between these models fails to establish a clear winner because it remains statistically insignificant.

4.2 Comparative Evaluation of Independent, Classifier Chain, and Ensemble Models

XGBoost served as the base learner to implement both One-vs-Rest (OvR) strategy and classifier chain approaches in light of its high individual success in prior assessments.

The ensemble of classifier chains produces the best or one of the best evaluation results while maintaining low performance variation across all metrics. The ensemble model delivers better accuracy scores as well as superior F1 weighted scores compared to each individual chain and surpasses the independent One-Vs-Rest classifier to highlight the benefits of handling interlabel dependencies. Chains 3, 6 and 10 present robust individual results but demonstrate higher level of variation compared to the combined model.

The ensemble method improves the model's identification of correct positive results better than standalone classification due to its enhanced sensitivity scores compared to independent assessments. The independent model showcases the most specific diagnostic outcomes although this performance comes with a disadvantage of missing important diagnostic cases. The ensemble achieves an equilibrium between high test accuracy and distinguished detection performance.

The Jaccard index demonstrates that ensemble predictions achieve maximum overlap compared to actual outcomes thus establishing the ensemble model as the best performer. This evaluation method produces the strongest matches because it effectively represents intricate diagnosis relationships that lead to reliable multi-class predictions.

The ensemble of classifier chains delivers the highest performance according to Figure 5 through its excellent ability to achieve balanced predictions and accuracy and robust results across five evaluation metrics. The results show that independent treatment of labels produces suboptimal outcomes during multi-label ECG classification. Multiple physician models will improve the

performance of medical signal analysis by reflecting interaction patterns between diagnosis labels while their ensemble demonstrates better stability qualities and clinical relevance for real-world implementations.

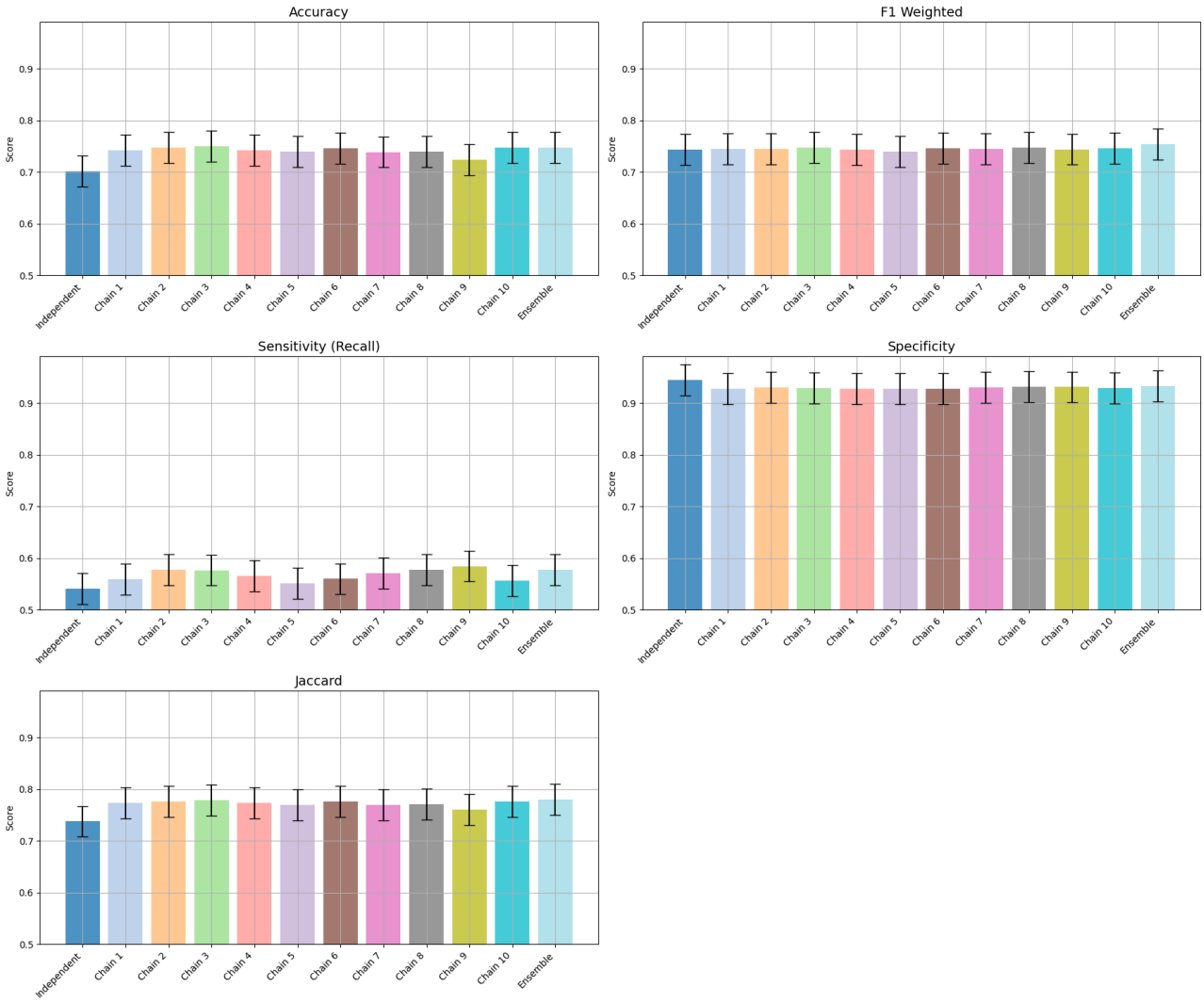


Figure 5. Comparative Performance of Independent, Classifier Chain, and Ensemble Models Across Five Evaluation Metrics

The figure 6 demonstrates ROC curves that show applications of ensemble classifier chain models with One-vs-Rest models among six ECG diagnostic classes. Both true positive rate and

false positive rate can be observed on the ROC curve while the area under the curve (AUC) provides a single number to measure overall class discrimination performance. Each of the two applied modeling approaches generates separate ROC curves in this multi-label setup to provide distinct insights about their abilities to classify individual labels.

The classification performance of both models reaches high standards across all six classes as their AUC values rise above 0.80 for every identified label. When analyzing Class 5 both OvR and ensemble models produce an outstanding AUC score of 0.97 indicating their exceptional capability to distinguish positive from negative instances in this category. Strong agreement exists between the models regarding identification of Class 0 and Class 4 conditions based on the excellent discrimination measured by AUC values of 0.93.

The ensemble classifier chain generates barely higher AUC values than OvR for Classes 1, 2 and 3 in the data sets. The ensemble classification method in Class 3 reaches an AUC value of 0.92 while OvR reaches only 0.91. Ensemble modeling provides minimal but consistent performance benefits in the diagnosis of Class 1 and Class 2 conditions. Even though these differences are minimal they indicate how modeling label dependencies by the ensemble chain results in better classification when ambiguity occurs.

Both approaches demonstrate consistent high performance without degraded results on any label and this observation demonstrates their strong operational stability. The ensemble classifier chain maintained high AUC scores throughout all classes and demonstrated better results in previous multi-metric assessments which establishes it as the optimal method for multilabel ECG classification.

The results from Figure 6 demonstrate that the ensemble classifier chain achieves superior strength as compared to alternative methods in diagnostic accuracy contexts affected by label dependencies. The validation tests strengthen the ensemble's ability to deliver consistent and

enhanced performance so it should be considered for implementation in clinical ECG systems.

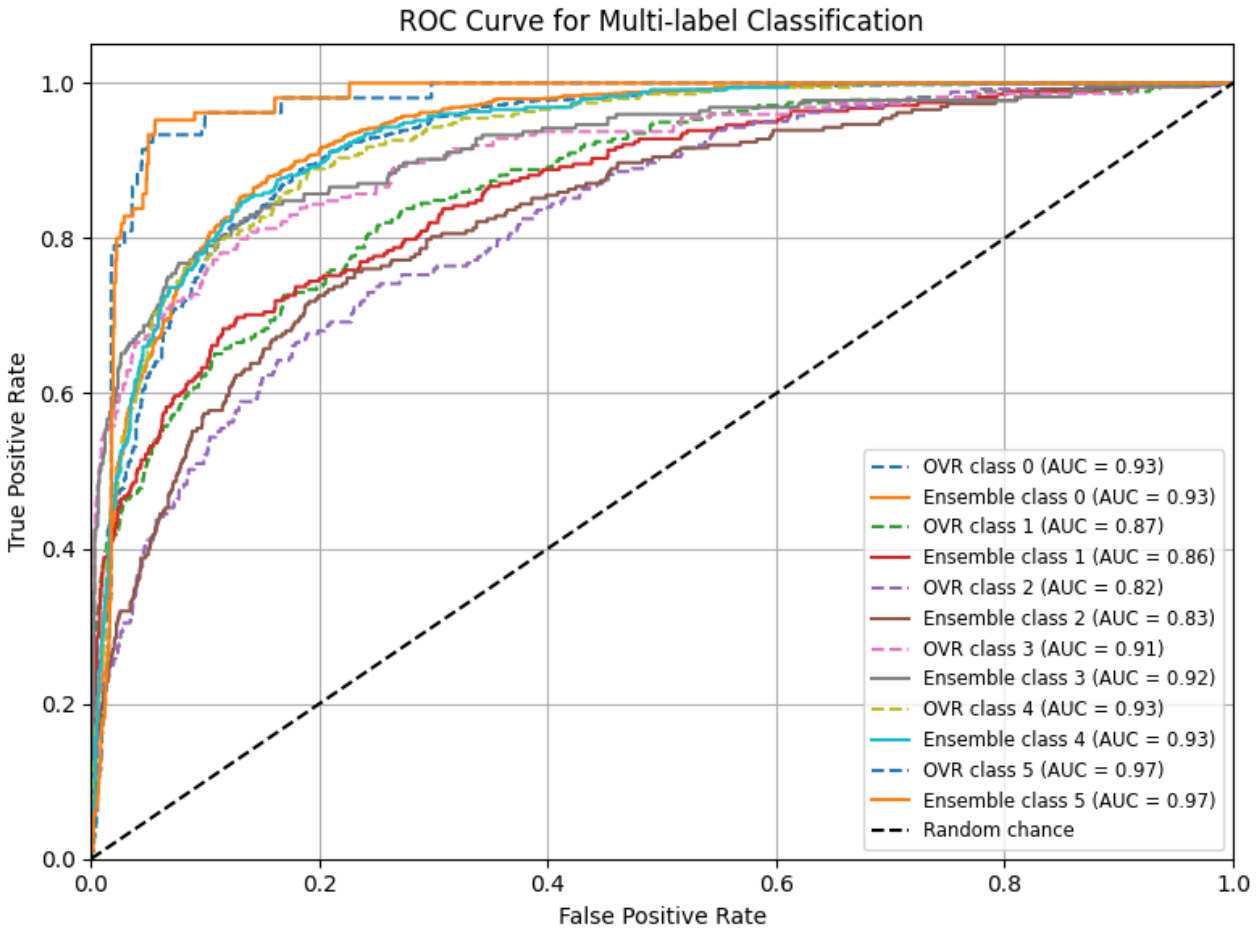


Figure 6. ROC Curve Comparison for One-vs-Rest and Ensemble Classifier Chain Models Across Multi-label ECG Classes

4.3 Feature Importance and Influence

The figure 7 shows a comparison of typical feature significance between a One-vs-Rest structure and a classifier chain ensemble which includes XGBoost as their base learner. The visualization shows the way different ECG-derived features obtain priority levels during multi-label classification operations for both models.

The supported features in the OvR model exist as distinct peaks featuring T Amplitude Mean along with RR Standard Deviation and ST-Segment T Mean among the most vital features. Ventricular repolarization amplitude stands out as the most influential feature of all in label prediction when executed independently according to T Amplitude Mean. The independent optimization of each binary classifier in OvR leads to its architectural design which does not consider between-label relationship effects.

As distinguished from the classifier chain ensemble the feature importance distribution across the different features is more spread out. The T Amplitude Mean and ST-Segment measures remain top priorities for the classifier chain ensemble yet the variation between its features remains minimized. The combination of chaining structure with label interdependence modeling allows the ensemble model to effectively use diverse input features when making predictions. The spread of feature importance across features makes the ensemble model robust and guards against overfitting thus leading to higher stable prediction performance as observed before.

The P Amplitude together with QRS Duration and R Amplitude proved essential for diagnosis in both models according to diagnostic results. The classifier chain ensemble appears to engage multiple features for decision-making through its holistic training process which results from label interactions.

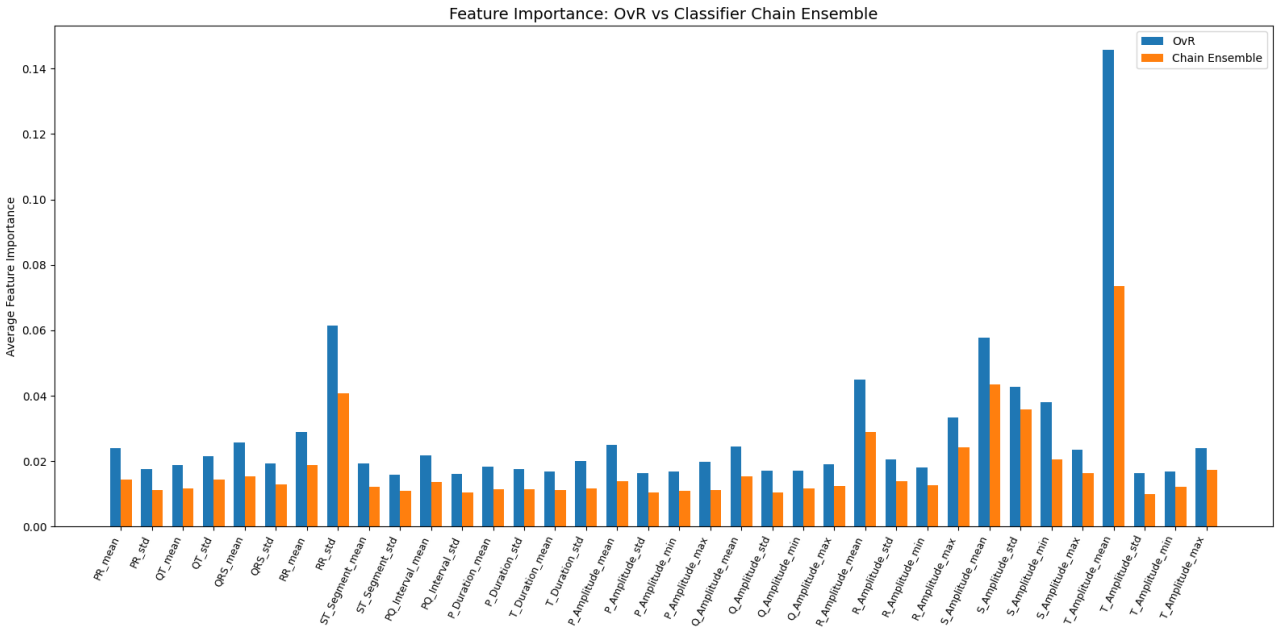


Figure 7. Average Feature Importance Comparison Between OvR and Classifier Chain Ensemble Models

The depicted figure shows average feature importance ratings from both One-vs-Rest and classifier chain ensemble models across every ECG-related feature. The OvR model allocates most importance weight to T Amplitude Mean and RR Standard Deviation yet the classifier chain ensemble distributes its weights across various features because of its multi-label learning approach.

4.4 Confusion Matrix and Error Analysis

The figure 8 displays confusion matrices for each diagnosis type among six major ECG diagnostic groups that include Normal (NORM), Myocardial Infarction (MI), Hypertrophy (HYP), Conduction Disturbance (CD), ST/T Change (STTC), and Abnormal (ABNORM). The extensive class-based tables enable medical interpretation of the model expectation outcomes by displaying both true positives (TP) and true negatives (TN) and false positives (FP) and false negatives (FN).

The NORM class receives exceptional performance from the classifier which generates 1,399 true positives together with 690 true negatives and 131 false negatives and 196 false positives. The model demonstrates excellent performance by accurately detecting normal ECGs alongside acceptable identification of incorrect cases.

A total of 116 true MI cases receives classification accuracy from the model yet 162 test cases remain undetected due to a higher number of false negative results. The model demonstrates caution when identifying myocardial infarctions possibly because various abnormal patterns appear similar to this condition.

The HYP classifier provides 68 correct detections together with 195 cases of undetected abnormalities. Hypertrophy cases show minimal waveform variations which make their patterns difficult to distinguish from typical or doubtful categories.

In CD cases (Conduction Disturbance) the classification system accurately forecasts 124 plus tests while producing 29 incorrect results indicating strong accuracy. Having 100 false negatives indicates that even though the tests display high specificity they still fail to detect multiple positive cases.

The model shows balanced results for STTC by identifying 213 true positives and misidentifying 140 false negatives and 82 false positives. The intermittence of ST/T abnormalities together with other diseases leads to this level of performance which medical teams find useful however there is room for improvement.

The classifier shows the greatest challenge in detecting ABNORM because its true positives (52) nearly match false negatives (53) and it produces 41 false positives. The model faces challenges classifying ABNORM patterns due to the diverse nonspecific abnormalities present in this category.

According to Figure 8 the model displays high success in normal ECG detection along

with good performance for CD and STTC yet it encounters difficulties identifying HYP and both ABNORM and MI kinds accurately. The model performance data can be used to guide future improvements which include calibrating the model per class and training it for costs and adding domain-specific rule implementations.

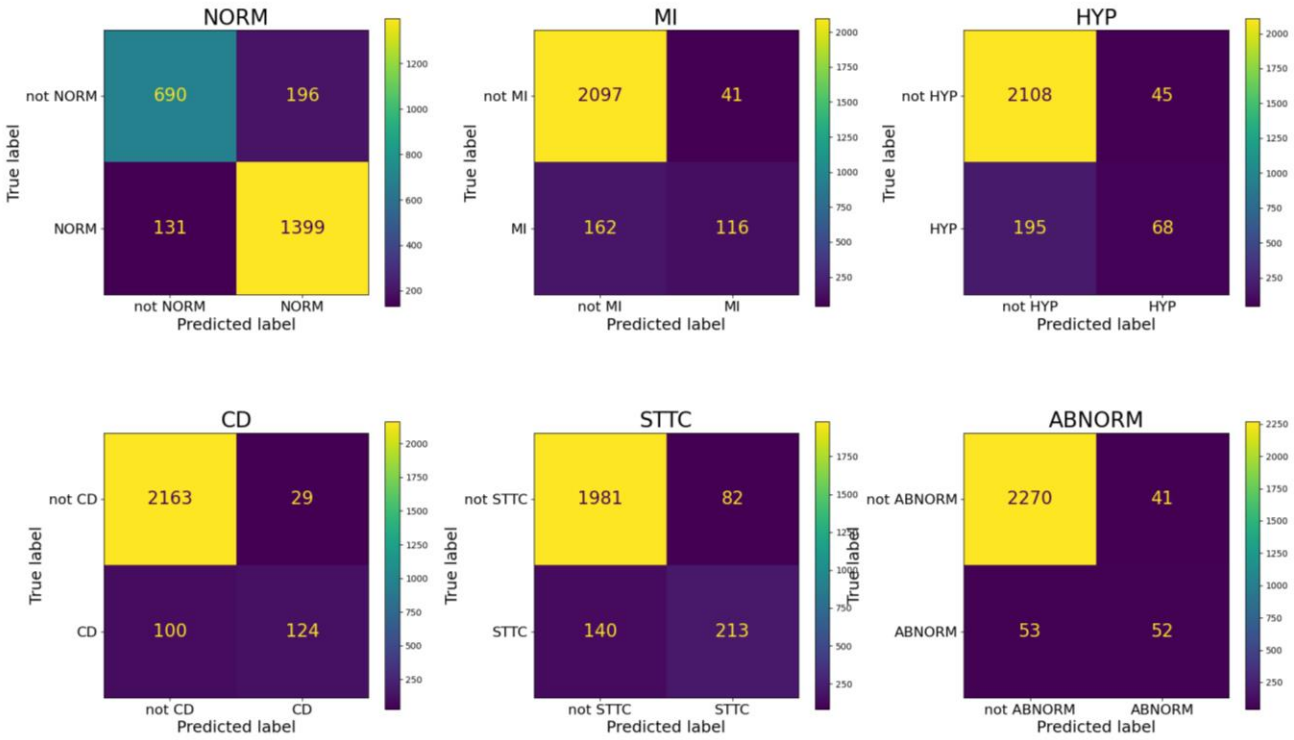


Figure 8. Per-Class Confusion Matrices for ECG Superclasses

Confusion matrices for each of the six ECG diagnostic superclasses: Normal (NORM), Myocardial Infarction (MI), Hypertrophy (HYP), Conduction Disturbance (CD), ST/T Change (STTC), and Abnormal (ABNORM). The different matrices reveal both actual and incorrect positive and negative detection rates to provide detailed insights about how the model performs within each class. Both NORM and CD diagnostic categories record high success rates from the classifier system yet the performance drops down when processing the additional problematic categories of HYP and ABNORM.

4.5 Hybrid Model Training Results

4.5.1 Deep Learning Model Training Results

The model tracking at training phase used accuracy and loss evaluation metrics for both datasets during 120 epochs. The figure 9 demonstrates training accuracy which started at 62% and rose to 75% through the entire training duration. The validation accuracy experienced an equal training pattern starting from 62% which leveled off at 72% following 120 epochs of training.

Training loss experienced a quick descent from its initial 0.5 levels to achieve a near 0.25 value by the conclusion of training which reflected efficient learning on the training data. The procedure started with a validation loss value of approximately 0.45 yet it exhibited steady decline up to its stabilization at 0.28 by the time training ended.

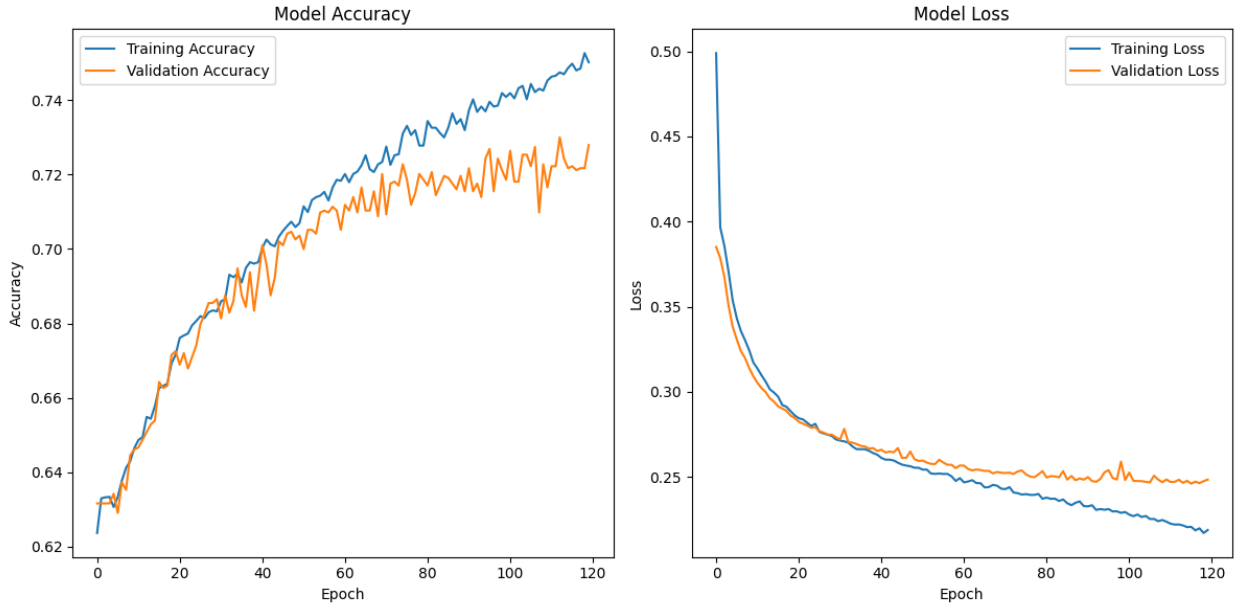


Figure 9. Training and Validation Accuracy and Loss Curves for Deep Learning Model

4.5.2 Performance of XGBoost Classifier Chain on Deep Learning Extracted Data

After extracting features from the deep learning model, an XGBoost model with a classifier chain was trained to perform multi-label classification on the ECG data. The model achieved the following performance across different classes, as shown in the table 4.

Table 4: Performance Metrics for Hybrid Classifier Chain Ensemble on Multi-Label ECG Classification

Class	n (Truth)	n (Classified)	Accuracy	Precision	Recall	F-Score
NORM	6116	6410	94.95	93.90	98.41	96.10
MI	1102	841	96.82	97.27	74.23	84.20
CD	1066	718	95.76	95.68	64.45	77.02
HYP	935	790	98.33	98.99	83.64	90.67
STTC	1463	1376	97.53	94.48	88.86	91.58
ABNORM	396	449	98.81	81.29	92.17	86.39

The overall accuracy of the model was 87.79%, with a weighted accuracy of 95.98%, indicating strong performance in detecting various cardiovascular conditions. The highest accuracy was achieved in the HYP and ABNORM classes, with accuracies above 98%, while MI and CD classes showed slightly lower performance, especially in terms of recall. The NORM class maintained good scores across all metrics, largely due to its high prevalence in the dataset and the class imbalance.

4.6 Hybrid Model Integration into the Web Platform

The hybrid model was integrated into the web platform so users could upload ECG images while getting diagnostic prediction probabilities. Users can submit single-lead ECG images to the platform that processes and transmits them to the hybrid model for assessment. The platform's interface visualizes user interaction with the model predictions as shown in Figures 10 and 11

alongside the image uploading process. The model's response time was measured from the moment the ECG image was uploaded to when the prediction was displayed, averaging around 22 seconds per instance.

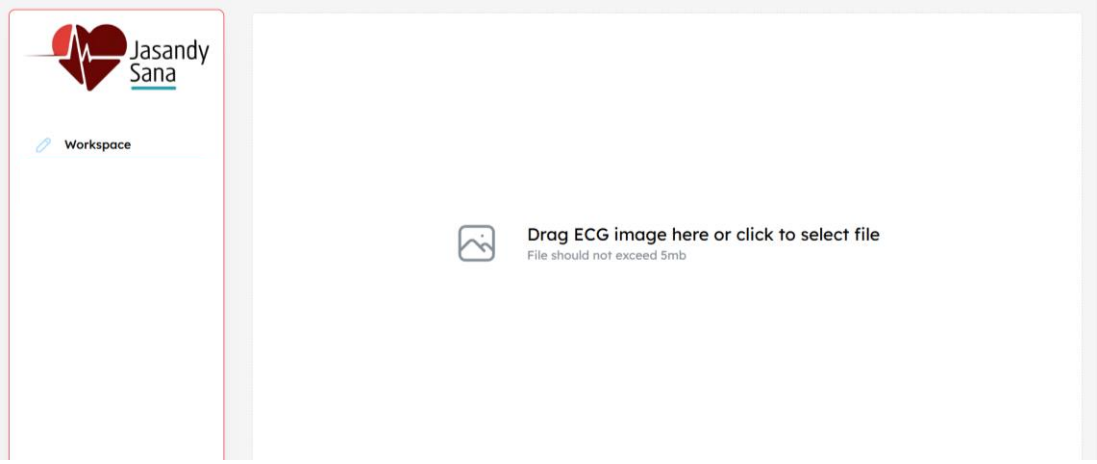


Figure 10. User Interface for ECG Image Upload

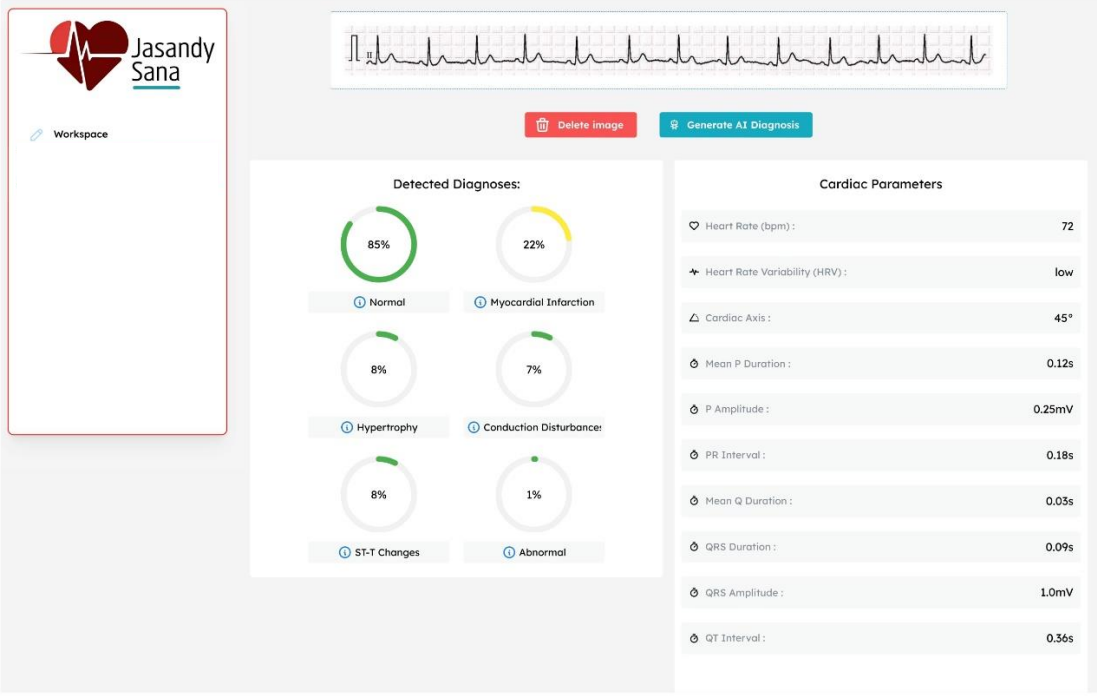


Figure 11. Model Prediction Results Display

4.7 Conclusion of Results

The performance benchmarking was conducted using conventional models that included LDA, Naïve Bayes, SVM, CART and k-Nearest KNN. The predictive models KNN and SVM produced the most accurate results with measurement outputs between 68% to 69%. Due to its fundamental feature independence presumption Naïve Bayes produced inaccurate results when working with ECG signal data that shows strong correlations between variables. Even though Naïve Bayes had lower precision, its excellent ROC AUC score indicated successful class ranking performance.

Random Forest and Gradient Boosting together with XGBoost and LightGBM and Voting and Stacking ensemble classifiers exceeded traditional model results in terms of performance. F1-score and ROC AUC values for ensemble methods established XGBoost and LightGBM and Stacking as excellent performers among these algorithms. These classifiers achieved accuracy levels ranging from 66% to 71%. The Voting classifier provided successful model performance in dealing with class imbalance situations. The models based on XGBoost and LightGBM exhibited the best Hamming loss results indicating they could predict multiple labels better than single models.

The deep learning model based on 1D CNN utilized ECG data directly for training which led to the evaluation of accuracy levels and loss measurements. A single test yielded 75% training precision and 72% validation precision while finally reaching training and validation loss of 0.25 and 0.28 respectively. The measured metrics show a strong ability for learning with effective generalization. The deep learning approach exhibited direct limitations when used separately against the classifier chain model for complex label interdependence detection. The CNN extracted ECG temporal patterns together with morphological characteristics successfully yet its basic

architecture prevented it from handling multi-label relationships which caused lower precision and recall specifically for rare classes MI and CD.

In contrast, the classifier chain model, built on XGBoost as a base learner, explicitly modeled label dependencies by incorporating previous label predictions as input features in a sequence of classifiers. This allowed the model to better handle co-occurring ECG abnormalities and improved performance in multi-label settings. The classifier chain achieved higher F1-scores and Jaccard index values compared to the deep learning model, indicating better alignment with true label sets across multiple conditions. Its architecture enabled more stable and balanced predictions, especially for overlapping or ambiguous diagnostic categories.

The hybrid model, which combined CNN-extracted features with the XGBoost classifier chain, yielded the best overall results. It achieved 87.79% overall accuracy and 95.98% weighted accuracy, outperforming both the deep learning and standalone classifier chain models. This confirmed that the fusion of deep feature representations with structured label modeling significantly enhances diagnostic performance in multi-label ECG classification.

The hybrid model was further integrated into a web platform, enabling users to upload ECG images and receive real-time diagnostic predictions. The system processed single-lead ECG images and returned results within approximately 22 seconds per image. This integration demonstrates the platform's potential for practical deployment in clinical environments, providing timely, automated support for cardiovascular diagnosis.

CHAPTER 5 - (DISCUSSION)

The main objective of this research was to create a machine learning ML model for ECG abnormalities classification which aimed at boosting diagnostic precision in medical environments. This approach used an ensemble model of deep learning and machine learning where CNN extracted features for processing multi-label ECG data. The research results show that the proposed model achieves superior performance beyond traditional models.

Subsequently the hybrid model succeeded in multi-label ECG classification through combining CNN-extracted features with classifier chain ensemble methods. The model delivered an 87.79% accuracy alongside a 95.98% weighted accuracy level thus outperforming traditional approaches because of their inadequate handling of complex labeling connections. The reported findings match previously documented research in automated ECG analysis because hybrid systems show superior performance through their combination of several feature extraction and classification techniques.

The classifier chain ensemble proved successful at modeling label interdependencies because it established models between myocardial infarction MI and conduction disturbances CD although these conditions commonly occur together. The new architecture delivered better prediction compared to One-vs-Rest since it captured relationships between different conditions during classification. Each class in the traditional method operated independently. The explicit modeling of relationships by the classifier chain generated better precision and recall while raising the overall F1-score mainly when dealing with rare conditions.

Several obstacles persist in using the hybrid model due to unbalanced class representation in the ECG dataset. The model experienced difficulty in distinguishing hypertrophy and some

abnormal patterns according to the provided error analysis and confusion matrices. Model detection of hypertrophy cases faced difficulties because these conditions demonstrate minimal waveform changes so the model needs specialized training or adjustments for effective detection of subtle heart abnormalities. The ABNORM class containing uncommon ECG abnormalities proved difficult for the model to identify correctly between its different conditions and normal or abnormal signals properly.

The training of the model was done by using the PTB-XL dataset yet this collection might not showcase the total range of ECG abnormalities across all population groups while HYP and ABNORM conditions exist in relatively low numbers within the dataset. The accuracy levels of these classes tend to decrease because of this problem. The practical single-lead ECG approach reduces the model's capacity to notice specific abnormalities that multi-lead ECGs detect more easily.

A temporal context addition to the model would help enhance performance because it provides better recognition of evolving abnormalities throughout time. Clinicians face difficulties understanding the reasons behind particular predictions made by the model because of its interpretability limitations.

The hybrid model demonstrates excellent capability for detecting ST/T changes together with conduction disturbances but could benefit from extra feature development or demographic and medical history inclusion to enhance detection accuracy in complex cases. The deep learning element enabled the 1D CNN to extract vital features from ECG signals but it was unable to properly address data relationships in multi-label contexts which led the idea to develop the classifier chain ensemble method for superior results.

A web platform integrating this model achieves essential progress in delivering ECG diagnosis automation which benefits clinical environments especially remote medical facilities.

The web platform enables medical personnel to deliver ECG images into the system for processing and diagnostic predictions that take about 22 seconds per image. The proactive deployment of this model demonstrates technical viability in healthcare facilities to provide fast heart disease diagnosis that improves medical results.

The practical deployment of this system needs additional evaluation and testing which must include expanded conditions combined with different demographic variables. The model's general applicability will be achieved through additional testing with diverse patient cohorts to perfect its functionality. The research should examine how to integrate additional advanced analytics techniques including deep learning attention mechanisms with the implementation of multiple ECG lead information that could boost diagnostic precision.

CHAPTER 6 - (CONCLUSION)

This research project focused on creating a strong automated ECG abnormality classification system to enhance both diagnostic precision and operational effectiveness for detecting heart tissue damage with plant enlargement and cardiac conduction problems as well as changes to ST/T wave patterns. Solution effectiveness was proven by the 87.79% overall accuracy and 95.98% weighted accuracy results yielded from using deep learning feature extraction with XgBoost classifier chains. The merged diagnostic approach handled multi-label classification issues of ECG diagnostic systems successfully because it recognized cardiovascular condition connections.

The web-based platform integration of this model enhanced its practical value because it enabled healthcare professionals in remote areas with limited resources to submit single-lead ECG images which received instant diagnostic results. The model shows practical potential to improve diagnostic processes in medical settings and promotes accessible automatic ECG assessment across healthcare facilities. Analysis is completed swiftly on the system through ECG images processing that takes approximately 22 seconds per image thus making it ideal for quick healthcare emergency responses.

Limitations remain despite promising outcomes from the study. The model experienced difficulties in recognizing rare abnormalities that included both hypertrophy and selected conditions from the ABNORM class because their electrocardiographic features were difficult to distinguish. The dataset being highly imbalanced, led to model overfitting, leaving a big question about model validation on the real clinic data. The comprehensive training dataset did not contain all ECG abnormality variations found between different population groups which could reduce the

model's ability to generalize across demographics.

Research into the model's performance enhancement for detecting unusual and complex abnormalities should concentrate on adding patient demographic characteristics and examining how additional ECG lead data affects classification results. Modern attention mechanisms should be applied to the model for better tracking of time-dependent abnormalities across periods.

REFERENCE LIST

[1] Elizabeth Wilkins *et al.*, "European Cardiovascular Disease Statistics 2017 edition," 2017. [Online]. Available: www.ehnheart.org

[2] N. Townsend *et al.*, "Epidemiology of cardiovascular disease in Europe," Feb. 01, 2022, *Nature Research*. doi: 10.1038/s41569-021-00607-3.

[3] Dale Dubin, *Rapid Interpretation of EKG's*. Cover Publishing Company, 2000.

[4] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern Recognit*, vol. 38, no. 1, pp. 133–142, Jan. 2005, doi: 10.1016/j.patcog.2004.05.014.

[5] S. Aziz, S. Ahmed, and M. S. Alouini, "ECG-based machine-learning algorithms for heartbeat classification," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-97118-5.

[6] M. AlGhatrif and J. Lindsay, "A brief review: history to understand fundamentals of electrocardiography," *J Community Hosp Intern Med Perspect*, vol. 2, no. 1, p. 14383, Jan. 2012, doi: 10.3402/jchimp.v2i1.14383.

[7] J. M. Di Diego, "ECG Waves and Signs: Ionic and Cellular Basis," 2020, pp. 117–148. doi: 10.1007/978-3-030-41967-7_5.

[8] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput Biol Med*, vol. 100, pp. 270–278, Sep. 2018, doi: 10.1016/j.compbiomed.2017.09.017.

[9] J. Zheng, G. Fu, K. Anderson, H. Chu, and C. Rakovski, "A 12-Lead ECG database to identify origins of idiopathic ventricular arrhythmia containing 334 patients," *Sci Data*, vol. 7, no. 1, p. 98, Mar. 2020, doi: 10.1038/s41597-020-0440-8.

[10] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Inf Sci Syst*, vol. 2, no. 1, p. 3, Dec. 2014, doi: 10.1186/2047-2501-2-3.

[11] J. He, S. L. Baxter, J. Xu, J. Xu, X. Zhou, and K. Zhang, "The practical implementation of artificial intelligence technologies in medicine," *Nat Med*, vol. 25, no. 1, pp. 30–36, Jan. 2019, doi: 10.1038/s41591-018-0307-0.

[12] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science (1979)*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.

[13] E. Alpaydın, *Machine Learning*. The MIT Press, 2021. doi: 10.7551/mitpress/13811.001.0001.

[14] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform," *Biomed Signal Process Control*, vol. 8, no. 5, pp. 437–448, Sep. 2013, doi: 10.1016/j.bspc.2013.01.005.

[15] Y. Sattar and L. Chhabra, *Electrocardiogram*. 2025.

[16] N. J. Ashley EA, *Cardiology Explained*. 2004.

[17] N. Frey and E. N. Olson, "Cardiac Hypertrophy: The Good, the Bad, and the Ugly," *Annu Rev Physiol*, vol. 65, no. 1, pp. 45–79, Mar. 2003, doi: 10.1146/annurev.physiol.65.092101.142243.

[18] K. Thygesen *et al.*, "Fourth Universal Definition of Myocardial Infarction (2018)," *J Am Coll Cardiol*, vol. 72, no. 18, pp. 2231–2264, Oct. 2018, doi: 10.1016/j.jacc.2018.08.1038.

[19] D. H. Smith, E. S. Johnson, M. L. Thorp, K. A. Crispell, X. Yang, and A. F. Petrik, "Integrating Clinical Trial Findings into Practice through Risk Stratification: The Case of Heart Failure Management," *Popul Health Manag*, vol. 13, no. 3, pp. 123–129, Jun. 2010, doi: 10.1089/pop.2009.0047.

[20] W. T. Harkness and M. Hicks, *Right Bundle Branch Block*. 2025.

[21] E. W. Hancock, B. J. Deal, D. M. Mirvis, P. Okin, P. Kligfield, and L. S. Gettes, "AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram," *J Am Coll Cardiol*, vol. 53, no. 11, pp. 992–1002, Mar. 2009, doi:

10.1016/j.jacc.2008.12.015.
[22] R. Wang, J. Fan, and Y. Li, “Deep Multi-Scale Fusion Neural Network for Multi-Class Arrhythmia Detection,” *IEEE J Biomed Health Inform*, vol. 24, no. 9, pp. 2461–2472, Sep. 2020, doi: 10.1109/JBHI.2020.2981526.
[23] A. Y. Hannun *et al.*, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nat Med*, vol. 25, no. 1, pp. 65–69, Jan. 2019, doi: 10.1038/s41591-018-0268-3.
[24] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, “Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL,” *IEEE J Biomed Health Inform*, vol. 25, no. 5, pp. 1519–1528, May 2021, doi: 10.1109/JBHI.2020.3022989.
[25] A. Y. Hannun *et al.*, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” *Nat Med*, vol. 25, no. 1, pp. 65–69, Jan. 2019, doi: 10.1038/s41591-018-0268-3.
[26] K. H. Le, H. H. Pham, T. BT. Nguyen, T. A. Nguyen, T. N. Thanh, and C. D. Do, “LightX3ECG: A Lightweight and eXplainable Deep Learning System for 3-lead Electrocardiogram Classification,” Jul. 2022.
[27] J. Li, S. Pang, F. Xu, P. Ji, S. Zhou, and M. Shu, “Two-dimensional ECG-based cardiac arrhythmia classification using DSE-ResNet,” *Sci Rep*, vol. 12, no. 1, p. 14485, Aug. 2022, doi: 10.1038/s41598-022-18664-0.
[28] Z. Yang *et al.*, “A coordinated adaptive multiscale enhanced spatio-temporal fusion network for multi-lead electrocardiogram arrhythmia detection,” *Sci Rep*, vol. 14, no. 1, p. 20828, Sep. 2024, doi: 10.1038/s41598-024-71700-z.
[29] P. Wagner *et al.*, “PTB-XL, a large publicly available electrocardiography dataset.,” *Sci Data*, vol. 7, no. 1, p. 154, May 2020, doi: 10.1038/s41597-020-0495-6.
[30] J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” *IEEE Trans Biomed Eng*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.
[31] F. Charte, A. Rivera, M. J. del Jesus, and F. Herrera, “Resampling Multilabel Datasets by Decoupling Highly Imbalanced Labels,” 2015, pp. 489–501. doi: 10.1007/978-3-319-19644-2_41.
[32] F. Charte, A. J. Rivera, M. J. del Jesus, and F. Herrera, “MLeNN: A First Approach to Heuristic Multilabel Undersampling,” 2014, pp. 1–9. doi: 10.1007/978-3-319-10840-7_1.