Knowledge-Infused Deep Learning for Enhanced Cardiovascular Diagnostics

Pharvesh Salman Choudhary

Knowledge-Infused Deep Learning for Enhanced Cardiovascular Diagnostics

A

Thesis submitted

for the award of the degree of

DOCTOR OF PHILOSOPHY

By

Pharvesh Salman Choudhary

DEPARTMENT OF ELECTRONICS AND ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781 039, ASSAM, INDIA

NOVEMBER 2025

Certificate

This is to certify that the thesis entitled "**Knowledge-Infused Deep Learning for Enhanced Cardiovascular Diagnostics**", submitted by **Pharvesh Salman Choudhary** (186102010), a research scholar in the *Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati*, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by him under our supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dated:
Guwahati.

Dr. Samarendra Dandapat
Professor
Dept. of Electronics and Electrical Engg.
Indian Institute of Technology Guwahati
Guwahati - 781039, Assam, India.

Acknowledgements

I would like to express my sincere and heartfelt gratitude to my supervisor, Prof. Samarendra Dandapat, for his constant support, encouragement, and invaluable mentorship throughout my Ph.D. journey. I am especially thankful for his belief in me and for the opportunity to be part of the Electro-medical and Speech Technology (EMST) research group at IIT Guwahati. His insightful feedback, inspiring approach to research, and continuous enthusiasm greatly shaped this thesis and my development as a researcher. He taught me the art of conducting research, providing both the freedom to explore novel ideas and the guidance needed to achieve my research goals.

I am sincerely thankful to my doctoral committee members, Prof. Prabin Kumar Bora, Prof. Rohit Sinha, and Dr. Rishikesh Kulkarni, for their continuous encouragement, constructive suggestions, and guidance in shaping the direction of my work. I also thank the Head of the Department of Electronics and Electrical Engineering, for providing the necessary infrastructure and a supportive research environment. I am grateful to the anonymous reviewers for their valuable feedback, which strengthened the quality of this work. My sincere thanks go to all faculty members of the department for their support and encouragement. I would like to especially thank Dr. L. N. Sharma, Senior Technical Officer, for his mentorship and continuous guidance throughout my research. I am also grateful to the technical and administrative staff for their cooperation during my time at IIT Guwahati. I extend my special acknowledgement to Mr. Bhriguraj Borah, Senior Technical Officer, Department of Computer Science and Engineering, for his support and timely access to computational facilities. Furthermore, I express my gratitude to the security, canteen, and maintenance staff for their assistance whenever required. I sincerely thank my M.Tech supervisor, Prof. Soumik Roy, for introducing me to the exciting field of biomedical signal processing at Tezpur University.

I would also like to thank my EMST labmates Dr. Mousumi, Himashree, Dr. Debasish, Ato, Omesh, Sumit, James, Akriti, Pooja and Sadhana for their camaraderie, support, and countless moments of help, both academic and personal. I am grateful to my friends on campus Pallab, Sudeshna, Dr. Anupama, Dollie, Dr. Madhusmita, Himadri, Dr. Atanu, Allen, Himanku, Riyaz, Raktim, and Nayan for making this journey more enjoyable with their companionship and moral support. I also extend my thanks to my seniors Dr. Tilendra Choudhary, Dr. Alex Paul Kamson, Dr. Vineeta Das, Dr. Eedara Prabhakararao, and Dr. Samarjeet and Dr Sibasis for their guidance and advice at various stages of my research.

Finally, my deepest and most heartfelt gratitude goes to my parents, sister, and all my family members for their constant blessings, encouragement, and silent prayers that sustained me throughout this journey.

Pharvesh Salman Choudhary

Abstract

Cardiovascular diseases remain a leading cause of mortality and morbidity on a global scale. Early and accurate diagnosis is crucial for effective treatment and improved patient outcomes. Differential diagnosis is a difficult task due to the complex nature of cardiac conditions, where multiple disorders can present similar symptoms and often co-exist. Electrocardiogram (ECG), a non-invasive tool capturing cardiac electrical activity, serves as the primary diagnostic modality in clinical practice. Manual ECG interpretation requires significant expertise, is time-consuming, and susceptible to variability, especially with subtle abnormalities or complex cases. This has motivated the development of automated ECG interpretation systems as clinical decision support tools, acting as an adjunct to the clinical decision process. Recent advances in deep learning have shown promising results in automated ECG analysis, several critical challenges persist in developing clinically reliable solutions.

Most existing deep learning approaches treat ECG analysis as a purely data-driven task, relying on abstract feature representations from deeper network layers to classify pathology, similar to general object classification tasks. However, in ECG interpretation, such approaches can overlook crucial pathological patterns that clinicians specifically examine for diagnosis. This thesis addresses this fundamental limitation, with the initial two works focusing particularly on myocardial infarction (MI) detection, a major contributor to cardiovascular mortality. Clinically, MI diagnosis relies on specific morphological alterations in ECG waveforms such as changes in QRS complex, ST-segment deviations, and T-wave abnormalities. To incorporate this crucial domain knowledge into the learning process, we develop the Cross-Task Attention Transfer (CTAT) framework, which leverages knowledge distillation between ECG delineation and MI detection tasks to guide the model's attention toward clinically relevant ECG characteristics. Extending this strategy of incorporating additional clinical context, in the second part of the thesis, we develop the Multi-task Physics-Constrained Hierarchical MI Classification framework (MTPCHi-MI). This framework uniquely combines physics-guided constraints with hierarchical classification, ensuring the model's predictions respect both the underlying electrophysiological

relationships between ECG leads and the natural diagnostic hierarchy in MI detection and localization. Comprehensive evaluation on multiple ECG databases demonstrates the superior performance of both frameworks compared to existing methods. Importantly, post-hoc interpretability analysis through Grad-CAM visualizations confirms that our models focus on clinically relevant ECG regions, providing transparent and diagnostically meaningful explanations for their predictions.

While accurate predictions are crucial for clinical diagnostic systems, equally important is their ability to express reliable confidence in these predictions. However, latest deep learning approaches provide only deterministic predictions without quantifying their uncertainty, potentially leading to overconfident decisions in ambiguous cases. In the third part of the thesis, we develop an Uncertainty-Aware ECG diagnostic framework for reliable multi-label cardiovascular abnormality detection. Our approach leverages variational inference to simultaneously achieve accurate classification, uncertainty quantification, and out-of-distribution detection. By incorporating a learnable condition interaction matrix, it captures interdependencies between cardiac abnormalities, enabling more coherent multi-label predictions. Extensive experiments on external databases demonstrate the framework's ability to reliably detect unfamiliar ECG patterns, providing an essential safety mechanism for clinical deployment.

Moreover, most automated ECG interpretation methods are designed for digital ECG signals and cannot be directly applied to ECG paper plots, which remain a major source of cardiac data, particularly in remote clinical settings and for accessing historical patient records. Finally, we present a novel multi-task learning framework for automated ECG image analysis. Our framework simultaneously addresses morphological cardiac ailment classification and cardiac rhythm classification while leveraging the structured layout of standard ECG prints. Through region-based feature extraction and task-specific analysis, the framework effectively mimics the clinical interpretation process, where experts systematically examine different ECG regions for comprehensive diagnosis. Extensive evaluation on both open-access and internal clinical ECG image databases demonstrates our framework's superior performance over existing methods, establishing an effective approach for automated interpretation of ECG paper records.

Keywords: Electrocardiogram, Knowledge Distillation, Multi-task Learning, Uncertainty Quantification, Hierarchical Classification, Variational Inference

Contents

List of Figures

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

List of Tables

List of Abbreviations

Acc	Accuracy
ACE	Average Component Entropy
ACCE	AccuRate Cardiocare Clinical ECG
Afib	Atrial Fibrillation
ASMI	Anterior Septal Myocardial Infarction
AUC	Area Under Curve
AV	Atrioventricular
AWGN	Additive White Gaussian Noise
BLSTM	Bidirectional Long Short-Term Memory
CAD	Coronary Artery Disease
CD	Conductive Disturbances
CNN	Convolutional Neural Networks
CPSC	China Physiological Signal Challenge
CVD	Cardiovascular Disorders
CTAT	Cross-Task Attention Transfer
DSC	Dice Similarity Coefficient
DTW	Dynamic time warping
ECG	Electrocardiography
ECHO	Echocardiography
ELBO	Evidence Lower Bound
EMD	Empirical Mode Decomposition
G12EC	Georgia 12-lead ECG Challenge
GAP	Global Average Pooling
GMM-VAE	Gaussian Mixture Model Variational Autoencoder
GRU	Gated Recurrent Unit

HD	Hausdorff distance
HYP	Hypertrophy
IMI	Inferior Myocardial Infarction
IOU	Intersection over union
KNN	K-Nearest Neighbors
KL	Kullback-Leibler
LA	Left Arm
LBBB	Left Bundle Branch Block
LDA	Linear Discriminant Analysis
LL	Left Leg
LN	Layer Normalization
LU	Lobachevsky University
LVH	Left Ventricular Hypertrophy
MI	Myocardial Infarction
MPI	Myocardial Perfusion Imaging
MRI	Magnetic Resonance Imaging
MTL	Multi-task Learning
NORM	Normal
NLL	Negative log likelihood
NLM	Non-local means
OA	Overall Accuracy
OOD	Out-of-Distribution
PPV	Positive Predictive Value
PR	Precision-Recall
PRD	Percentage root-mean-square difference
PTB-XL	Physikalisch Technische Bundesanstalt Extra Large
RA	Right Arm
RBBB	Right Bundle Branch Block
RDN	Region Delineation Network
ROC	Receiver Operating Characteristic
RNN	Recurrent Neural Networks

RVH	Right Ventricular Hypertrophy
SA	Sinoatrial
SARRH	Sinus Arrhythmia
SBRAD	Sinus Bradycardia
Sen	Sensitivity
Spe	Specificity
SPH	Shandong Provincial Hospital
SR	Sinus Rhythm
SSIM	Structural similarity index
STACH	Sinus Tachycardia
STTC	ST/T Changes
SVM	Support Vector Machines
TIE	Tree Induced Error
TOL	Tolerance
VMD	Variational Mode Decomposition
WEDD	Wavelet Energy based Diagnostic Distortion
WHO	World Health Organization

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

1

Introduction

Contents

The landscape of cardiovascular diagnostics stands at a pivotal crossroads, where conventional diagnostic methodology meet the precision of automation. Cardiovascular disorders (CVDs) refer to a wide range of health conditions that affect the heart and/or blood circulation system. In clinical practice, assessment of these cardiovascular conditions require systematic analysis of temporal cardiac electrical patterns along with diverse physiological indicators, which forms the basis for comprehensive diagnostic evaluation and effective patient care [1]. Disease diagnosis is a complex task that often requires the analysis of different modalities of patient data. This makes the diagnostic process particularly challenging, as pathologies can manifest through diverse physiological and biochemical changes [2,3]. Conventional diagnostic approaches, though foundational, are increasingly augmented by advanced computational systems that can detect subtle patterns which can often escape routine clinical observation. In recent years, advanced automated systems have emerged in cardiovascular diagnostics, some of which have even reached cardiologist-level precision [4]. The goal of such systems is not to replace clinical judgment but to provide clinicians with deeper insights into disease patterns and support more comprehensive diagnostic assessments.

CVDs are responsible for most number of deaths and health complications on a global scale, and their burden continues to rise, driven by the increasing prevalence of risk factors such as obesity, diabetes mellitus, hypertension, and unhealthy lifestyle choices [5]. According to the World Health Organization (WHO), an estimated 17.9 million deaths annually are due to CVDs, with the impact being more pronounced in low- and middle-income nations. The major contributors to these fatalities include disorders like coronary artery disease (CAD), which affects the blood vessels supplying the heart and can lead to myocardial infarction (MI); conductive disorders, in which the heart's electrophysiology is disrupted, causing abnormal heart rhythms such as atrial fibrillation; hypertrophic cardiomyopathy, which involves thickening of the heart muscle, impairing its function; and heart failure, a condition in which the heart is unable to pump blood effectively, often as a result of other CVDs [6]. Many of these disorders are progressive in nature, gradually worsening over time, and can lead to severe complications and even mortality without timely treatment [7]. Therefore, precise diagnosis and timely treatment initiation is crucial for patients with these conditions.

In clinical practice, clinicians employ a combination of diagnostic tools or modalities to accurately evaluate cardiovascular health [8]. These tools include electrocardiography (ECG), echocardiography (ECHO), coronary angiography, myocardial perfusion imaging (MPI), and cardiac magnetic resonance imaging (MRI), each providing a unique perspective into the structure and function of the heart. The choice of the appropriate diagnostic tool depends on the patient's specific symptoms, risk factors, and overall clinical presentation. Generally, a combination of these tools are employed to provide a comprehensive assessment of cardiovascular health and guide a suitable treatment strategy. Among all these diagnostic

modalities, ECG, being rapid, cost-effective and non-invasive, is widely used as a primary screening tool for assessing cardiovascular health [9]. It measures the heart's electrical activity, recorded noninvasively by placing electrodes at specific body sites. Interpreting ECG is a non-trivial task, given the complex nature of cardiac pathologies and their variable electrical manifestations. The accurate interpretation of ECG readings requires extensive knowledge and experience, as subtle changes in the waveforms can indicate a wide range of cardiovascular conditions. Even for skilled clinicians, analyzing ECG data can be time-consuming and challenging, particularly in emergency situations where rapid diagnosis is critical [10] [11,12]. Moreover, with the increasing prevalence of CVDs worldwide, the need for accurate ECG interpretation has grown significantly. The vast volume of ECG data generated daily poses a significant challenge to healthcare systems, often resulting in delays in diagnosis and treatment. Therefore, automated ECG interpretation is of critical importance as a support system for clinicians, enabling faster and more objective clinical decision-making.

To address these challenges, there has been a growing interest in developing automated diagnostic systems for ECG interpretation. However, current approaches often adopt a purely data-driven approach [13, 14], disregarding any domain knowledge integration. The performance of such an approach is heavily dependent on the quality of the training data [15]. Also, most of the current methods do not handle the uncertainty in the prediction and struggle with generalization when encountering data distributions or disease patterns outside their training distributions [16]. These models are often overconfident in predictions, producing predictions with high confidence even when the outputs are incorrect. Ideally, models should abstain from predictions when prediction uncertainty is high. For medical AI systems, it is crucial that models provide confidence scores alongside their predictions to aid clinical decision-making. Additionally, the majority of automated ECG interpretation methods focus primarily on digital ECG signals, with limited attention given to ECG plot images, despite their prevalence in clinical settings. Developing robust methods to effectively process and interpret these ECG plot images is crucial for ensuring the widespread applicability of automated diagnostic systems across diverse healthcare environments.

This dissertation aims to advance automated CVD detection and classification by developing diagnostic frameworks that leverage medical domain knowledge, incorporate uncertainty modeling for ECG-based systems, and develop methods for precise interpretation of ECG plot images. This chapter begins with a concise overview of cardiovascular physiology, followed by an introduction to electrocardiography and its importance in clinical practice. We then review existing automated methods for ECG-based CVD detection. Following this, we discuss the motivation behind our research. The chapter concludes by presenting the thesis contributions and outlining its structure.

1.1 Fundamentals of Cardiovascular Physiology

The human cardiovascular system comprises a complex network of the heart and blood vessels and is responsible for transporting oxygen, nutrients, hormones, and metabolic waste products throughout the body, supporting the function of nearly every organ system. At the center of it is the heart, a rhythmically beating muscular organ that serves as the central pump driving blood circulation throughout the body. It operates through two interconnected circuits: the pulmonary and systemic circulations, driven by the heart. The operation of the heart is further divided into a cardiac cycle consisting of systole (contraction phase) and diastole (relaxation phase). During systole, the heart's ventricles contract, ejecting blood into the aorta and pulmonary artery, while during diastole, they relax, allowing for filling with blood from the atria. In pulmonary circulation, the right ventricle pumps deoxygenated blood to the lungs through the pulmonary arteries. In the lungs, carbon dioxide is exchanged for oxygen, and the oxygen-rich blood returns to the left atrium through the pulmonary veins. Systemic circulation begins as the left ventricle pumps oxygenated blood through the aorta to the entire body [17]. As the tissues use oxygen and nutrients, the deoxygenated blood carrying carbon dioxide and waste returns to the right atrium via veins. The heart serves as the central pump, while arteries, veins, and capillaries ensure the efficient delivery and exchange of substances. Figure 1.1 illustrates this dynamic process, showing the heart, lungs, and body with the flow of oxygenated and deoxygenated blood.

The heart's rhythm is maintained by an electrical conduction system, beginning with the sinoatrial (SA) node, often called the heart's "natural pacemaker." This electrical impulse spreads through the atria to the atrioventricular (AV) node and continues down specialized fibers, ensuring coordinated contraction of heart chambers. Any disruption in this conduction pathway can lead to arrhythmias (irregular heartbeats), that may impair the heart's ability to pump blood effectively. The overall performance of the heart can be quantified by the cardiac output, which is evaluated as a product of heart rate (beats per minute) and stroke volume (amount of blood ejected per beat). The heart's function depends on its muscular structure and the blood supply to the the heart tissue itself (coronary circulation). Blockages in the coronary arteries can damage the heart tissue and compromise its pumping efficiency.

Given the critical role of the heart in maintaining overall health, impairments can arise from multiple factors, such as issues with the conduction system, myocardium, coronary arteries, heart valves, or pericardium. Disorders affecting these components can significantly impact the heart's performance and lead to various CVDs. For instance, conduction disturbances, such as bundle branch blocks or atrioventricular (AV) blocks, disrupt the heart's electrical impulse transmission, leading to abnormal heart

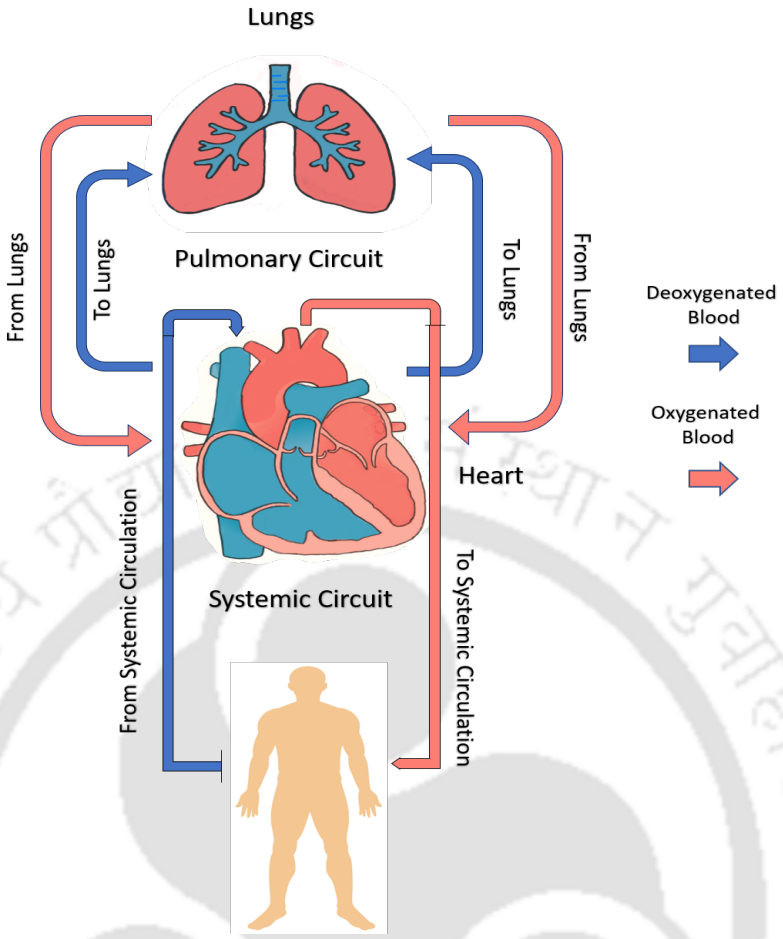

Figure 1.1: Illustration of the human cardiovascular system, highlighting the heart's central role in pumping oxygenated blood to the body and returning deoxygenated blood to the lungs for oxygenation.

rhythms. Another condition, hypertrophic cardiomyopathy, affects the heart muscle itself, in which the myocardium becomes thickened, particularly in the left ventricle, affecting the pumping efficiency. One of the most common CVDs is coronary artery disease, which occurs due to the buildup of atherosclerotic plaques in the coronary arteries, leading to reduced blood flow to the heart muscle. When the heart tissue is deprived of oxygen and nutrients, it can result in myocardial ischemia and, eventually, myocardial infarction if left untreated. It causes a portion of the myocardium to die, which can be permanent without timely treatment. Therefore, accurate diagnosis and timely treatment are crucial for appropriate CVD management. One of the primary and early diagnostic tools for assessing heart health is the ECG. It records the heart's electrical activity through electrodes placed on the skin's surface. It provides information about the heart's rhythm, conduction efficiency, and any potential abnormalities indicative of underlying CVDs. The following section will describe the principles of ECG and its role in diagnosing various CVDs, highlighting its importance in the early detection and management of CVDs.

1.2 Electrocardiography and its Clinical Significance

The heart's ability to pump blood effectively is governed by the intricate interplay of electrical impulses that originate and propagate throughout its specialized conduction system. The cardiac conduction system includes the sinoatrial node, atrioventricular node, and His-Purkinje system, which are responsible for the generation and propagation of electrical impulses that trigger the coordinated contraction of the heart muscles [18]. This electrical activity can be recorded and assessed using ECG, which represents the heart's electrical activity as a series of waveforms (P wave, QRS complex, and T wave) that correspond to different phases of the cardiac cycle [19], as shown in Figure 1.2. In Figure 1.3 we illustrate the cardiac electrical activity and its corresponding mechanical events. The P wave represents atrial depolarization, or the electrical activation of the atria, initiated by impulses from the sinoatrial (SA) node. This wave signals the contraction of the atria, allowing blood to flow into the ventricles. Following the P wave, the QRS complex appears and corresponds to ventricular depolarization, the rapid electrical activation of the ventricles. This complex is typically larger than the P wave due to the greater muscle mass of the ventricles, and it triggers their contraction, pumping blood into the pulmonary and systemic circulation. Finally, the T wave reflects ventricular repolarization or the process of restoring the ventricles' resting state in preparation for the next cycle. Together, these waveforms provide insights into the heart's electrical and mechanical function, enabling the diagnosis of a wide range of cardiovascular conditions.

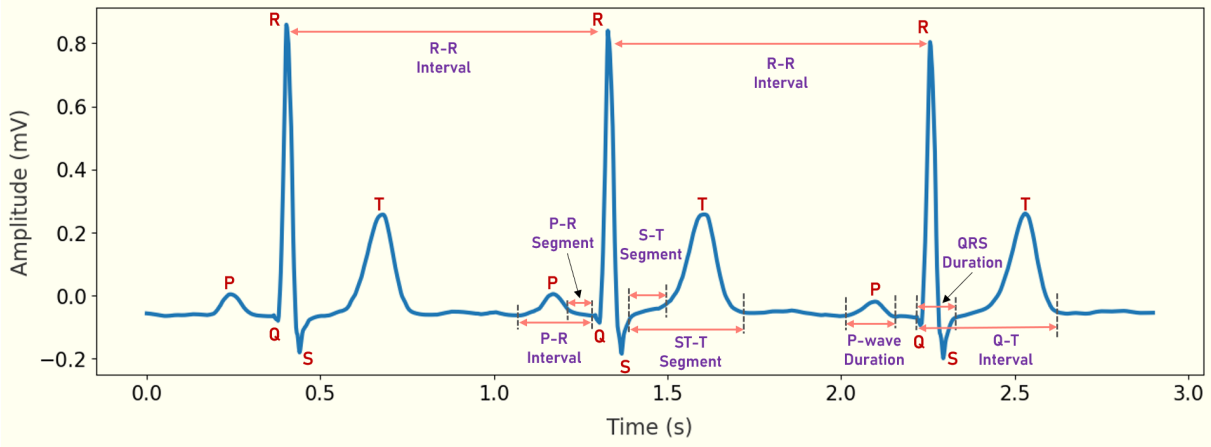

Figure 1.2: Standard ECG waveform illustrating three cardiac cycles with labeled components. Key features include P waves, QRS complexes, and T waves, along with clinically significant intervals (R-R, P-R, QRS duration, S-T segment, and Q-T interval) that serve as the foundation for automated ECG analysis and classification.

In clinical settings, the 12-lead setup is most widely used forms of ECG acquisition. This setup involves placing ten electrodes on the patient's body, four on the limbs and six on the chest [20]. The limb electrodes are positioned on the right arm (RA), left arm (LA), right leg (RL, used as a neutral reference), and left

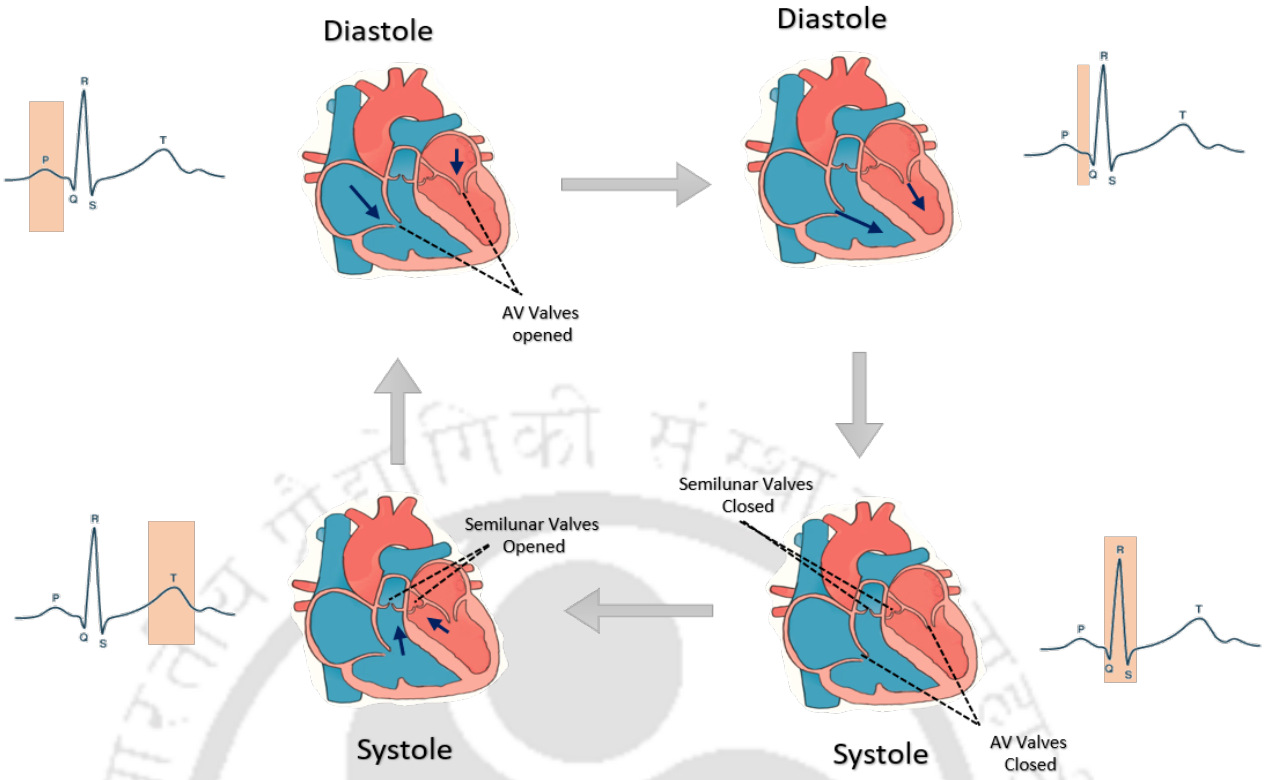

Figure 1.3: Cardiac cycle phases with corresponding ECG waves. The diagram shows the sequential relationship between heart valve movements during diastole and systole (AV and semilunar valves) and their associated ECG waveforms (P, QRS, and T waves). Arrows indicate the direction of cycle progression.

leg (LL). The six chest electrodes (V1 to V6) are placed at specific positions around the chest, capturing electrical signals from the anterior, lateral, and posterior aspects of the heart. The 12 leads derived from these electrodes are divided into six limb leads (I, II, III, aVR, aVL, and aVF) and six precordial (chest) leads (V1-V6). The limb leads provide views of the heart's electrical activity in the frontal plane, while the chest leads offer insights into its electrical patterns in the horizontal plane. The 12-lead ECG is the gold standard for comprehensive cardiac evaluation, but it is limited to short-term recordings in clinical settings. For long-term monitoring, ECG recording setups such as Holter or portable ECG monitors, including patch-based ECG devices, are often employed by clinicians to characterize occasionally occurring abnormal events that may be missed during routine hospital visits. These wearable devices generally record single-lead ECG continuously, typically for 24 to 48 hours or even longer, depending on the specific device and clinical requirements.

The standard 12-lead ECG configuration thus provides a comprehensive view of cardiac electrical activity from different spatial perspectives, enabling detailed assessment of cardiac function. The morphology of these electrical patterns can be significantly altered by various cardiovascular conditions. In this work, we

focus on several major CVDs and their characteristic ECG manifestations such as myocardial infarction (MI), hypertrophy (HYP), conductive disturbances (CD), and Atrial Fibrillation (Afib).

- **Myocardial Infarction**

  MI is a coronary artery disease resulting from interruption of blood flow to the myocardium, leading to myocardial ischemia and subsequent cell death [21]. It is a complex and multifactorial disorder caused due to rupture of atherosclerotic plaque and the formation of unstable plaques in the coronary artery [17]. This ischemic cascade triggers a series of biochemical and cellular changes, ultimately resulting in cardiomyocyte necrosis. The alteration in ECG morphology during MI dynamically evolves over time. The early manifestation of MI on ECG morphology involves the emergence of hyperacute T waves, usually noticeable within minutes of the onset of coronary occlusion. Following the hyperacute T wave, ST-segment elevation typically occurs, usually within 30 minutes of the onset of coronary occlusion [22]. ST-segment elevation is most pronounced in the leads that correspond to the area of the myocardium that is being infarcted. Over time, the ST-segment elevation will resolve, and T-wave inversion will develop. The Q-waves will typically appear within 12 hours of the onset of MI, and they may persist indefinitely [23–25]. Figure 1.4 illustrates a heart with an occluded coronary artery and evolving alterations in ECG waveform morphology linked to MI.

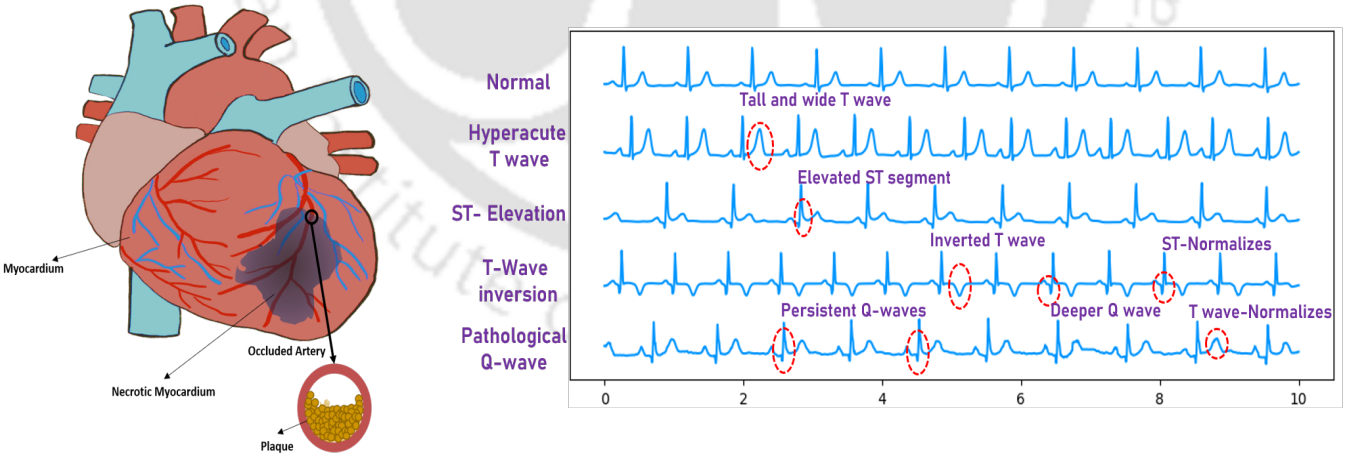

Figure 1.4: Illustration of heart exhibiting a blocked coronary artery and corresponding electrocardiogram alterations linked to MI

- **Hypertrophy**

  Cardiac hypertrophy refers to the abnormal enlargement of the heart muscle due to increased work-load, characterized by an increase in cardiomyocyte size. This structural adaptation typically occurs

in response to sustained hemodynamic stress, manifesting as increased ventricular wall thickness or chamber enlargement [26]. At the cellular level, this involves complex signaling pathways that trigger increased protein synthesis and sarcomere organization, leading to enlarged cardiomyocytes. The condition primarily manifests as either Left Ventricular Hypertrophy (LVH) or Right Ventricular Hypertrophy (RVH). LVH, most commonly resulting from systemic hypertension or aortic valve disease, is characterized by left ventricular wall thickening and produces distinct ECG patterns including increased QRS amplitudes in left-sided leads, deep S waves in right precordial leads, left axis deviation, and secondary ST-T wave changes indicating strain pattern [27]. Conversely, RVH, often associated with pulmonary hypertension or congenital heart defects, results in right ventricular wall thickening and manifests in ECG as dominant R waves in right precordial leads, deep S waves in left-sided leads, right axis deviation, and T wave inversions in right precordial leads [28] as shown in Figure 1.5. These structural alterations in cardiac architecture produce characteristic ECG patterns that serve as crucial diagnostic markers for identifying and monitoring the progression of cardiac hypertrophy.

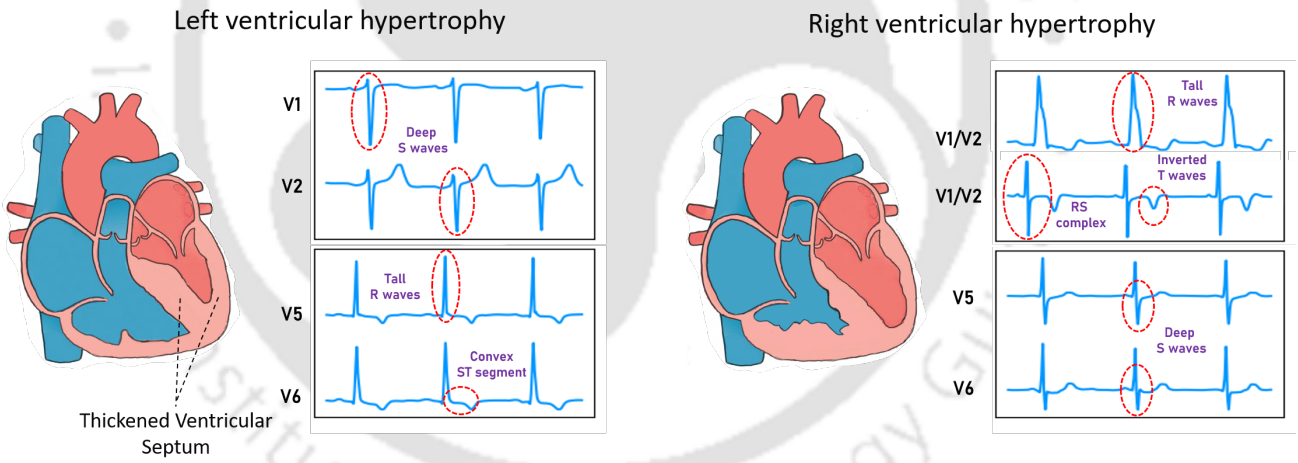

Figure 1.5: Anatomical and electrocardiographic characteristics of cardiac hypertrophy. Left: ventricular hypertrophy (LVH) with thickened left ventricular wall and corresponding ECG changes. Right: ventricular hypertrophy (RVH) showing right ventricular wall thickening and associated ECG patterns.

- **Conductive Disorders**

  Cardiac conductive disturbances encompass a group of disorders affecting the heart's electrical conduction system, which can occur at various anatomical locations from the sinoatrial node to the Purkinje fibers. These abnormalities disrupt the normal propagation of electrical impulses through the heart, leading to specific patterns of conduction delay or block. The main categories include Sinoatrial (SA) node dysfunction, Atrioventricular (AV) blocks, and bundle branch blocks. AV blocks,

characterized by impaired conduction between atria and ventricles are classified into three degrees: first-degree showing prolonged PR interval, second-degree exhibiting intermittent failure of conduction, and third-degree displaying complete dissociation between atrial and ventricular activity [29]. Bundle branch blocks occur when conduction is impaired in either the left or right bundle branches. Left Bundle Branch Block (LBBB) manifests as widened QRS complexes with characteristic morphology in left-sided leads, while Right Bundle Branch Block (RBBB) shows distinctive patterns in right-sided leads, both typically resulting in QRS duration exceeding 120ms [30, 31], as demonstrated in Figure 1.6. These conduction abnormalities produce specific ECG patterns that are crucial for diagnosis, risk stratification, and therapeutic decision-making.

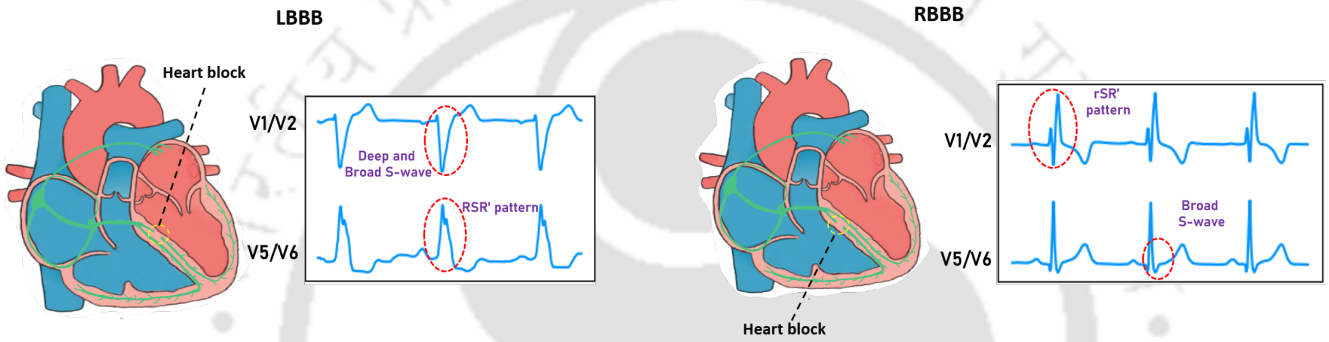

Figure 1.6: Schematic representation of bundle branch blocks and their ECG manifestations. Left: Anatomical pathway and ECG changes in left bundle branch block (LBBB) with characteristic QRS widening and altered ventricular activation sequence. Right: RBBB showing conduction abnormality and distinctive RSR' pattern in right precordial leads.

- **Atrial Fibrillation**

  AF represents a complex cardiac arrhythmia characterized by uncoordinated electrical activation of the atria. The underlying mechanism involves multiple chaotic electrical circuits or ectopic foci, primarily originating from the pulmonary veins, disrupting normal cardiac rhythm. These irregular impulses cause the atria to quiver rather than contract effectively [32]. The electrocardiographic manifestations are distinctive: normal P waves are absent and replaced by rapid, irregular fibrillatory waves (f-waves) with frequencies between 350-600 beats per minute, most evident in leads V1 and II [33]. As these chaotic electrical impulses reach the AV node, they result in irregular conduction to the ventricles, producing the characteristic 'irregularly irregular' rhythm seen in the QRS complexes. These pathophysiological changes and corresponding ECG manifestations are illustrated in Figure 1.7, which demonstrates the chaotic atrial activity and resulting loss of organized P waves. This disruption in normal cardiac electrical activity not only affects mechanical function but also increases

susceptibility to thromboembolic complications.

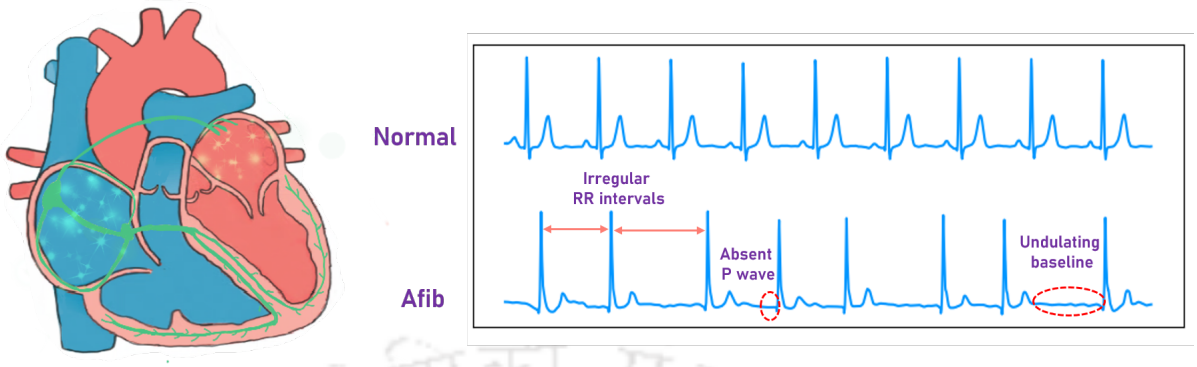

Figure 1.7: illustration of pathophysiology of atrial fibrillation. Left: Cross-sectional anatomy of heart showing multiple chaotic electrical circuits in atria. Right: Corresponding ECG manifestations showing replacement of P waves with fibrillatory waves and irregular ventricular response.

1.3 Automated Diagnostic of Cardiovascular Disorder -A Review

There is an extensive body of research on automated clinical predictions using ECG, with early computerized interpretation systems already established in the 1990s [34–36]. Building on this foundation, modern methods can broadly be classified into two categories: Conventional feature-engineering based methods [37–40] and deep learning methods [41–44]. Feature-engineering based methods involve the extraction of handcrafted features from the ECG signal, which are then used as inputs for traditional machine learning algorithms. These methods rely on expert knowledge to identify and compute relevant features that can diagnose different cardiac conditions. Common classification algorithms like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forests, Decision Trees, Linear Discriminant Analysis (LDA), and Artificial Neural Networks (ANN) have been extensively used in literature to evaluate these engineered features. Though these classical machine learning approaches have shown promise in ECG analysis, their performance heavily depends on the quality and comprehensiveness of the extracted features. Feature-engineering based methods offer interpretability and computational efficiency, with features that often have clear clinical meanings. However, these methods are fundamentally limited by their reliance on predefined features, which may not capture the full complexity of ECG signals and can miss subtle patterns crucial for accurate diagnosis. The manual feature extraction process is also time-consuming and requires extensive domain expertise, potentially introducing human bias into the analysis. On the other hand, deep learning methods, in recent years have emerged as a powerful alternative, capable of learning hierarchical representations of the ECG data without the need for manual feature extraction. Deep learning methods address the limitations of traditional approaches by automatically learning optimal representations directly

from raw ECG data. This capability allows them to discover complex, non-linear patterns and subtle signal characteristics that may not be apparent to human experts. While deep learning approaches require larger datasets and more computational resources, and often lack the interpretability of traditional methods, their superior performance in capturing intricate temporal and morphological patterns in ECG signals has made them increasingly attractive for clinical applications. Both of these approaches have been applied to a wide range of ECG classification and prediction problems. In the following subsections, we explore some notable works under these approaches.

1.3.1 Feature-Engineering methods in Cardiovascular Diagnosis

The early works for the ECG based automated detection of CVDs predominantly relied on feature engineering methods [14, 45, 46]. These approaches involved manually extracting meaningful features from ECG data to represent patterns indicative of heart conditions. The process of feature extraction heavily relied on domain knowledge and signal processing techniques to identify and quantify relevant information from the ECG waveform. The extracted features are then used as inputs for traditional machine learning for doing the classification. These features can be derived from various aspects of the ECG waveform, such as temporal, morphological, and frequency-domain characteristics, each offering unique insights into the nature of ECG signals. Temporal features capture timing-based characteristics of the ECG signal, such as intervals between successive heartbeats, peak-to-peak distances, and overall heart rate variability. Morphological features refer to the shape and amplitude characteristics of specific ECG waveforms, including the P wave, QRS complex, and T wave. These features often provide critical information about cardiac structure and electrical activity, enabling detection of myocardial infarctions, hypertrophy, and other structural heart diseases. For instance, chazal et al. [47] in their study for classification of heartbeats employed temporal features such as QRS-duration, T-wave duration, pre and post R-R intervals along with eight groups of morphological features derieved from different regions of the heartbeat. Similarly, Arif et al. [48] examined beat-specific temporal features such as T-wave amplitude, Q-wave characteristics, and ST level deviation for MI detection and localization. Romero et al. [49], in their assessment of myocardial ischemia, investigated morphological alterations during ischemic events by analyzing three distinct angles derived from R-wave slopes within the QRS complex. Maji et al. [45] focused on temporal and morphological characteristics surrounding ventricular activity region (QRS complex and T wave) of ECG for multiple arrhythmia detection. While temporal and morphological features provide valuable diagnostic insights, they present inherent limitations in capturing subtle cardiac abnormalities. Additionally, the complex manifestation of certain cardiac pathologies may extend beyond what traditional morphological analysis can capture. The time-varying

nature of cardiac electrical activity and the presence of overlapping waveform patterns further complicate the extraction of reliable diagnostic markers.

Frequency-domain methods break down ECG signals into fundamental frequency components, uncovering rhythmic patterns and subtle variations that might remain undetected in time-domain examination. Researchers have explored multiple decomposition approaches, including Fourier analysis, Wavelet transforms [38, 39], Empirical Mode Decomposition (EMD) [50], and Variational Mode Decomposition (VMD) [51–53]. Each approach offers unique insights into the spectral characteristics of cardiac signals. Sharma et al. [38] applied wavelet decomposition to 12-lead ECG signals, separating clinical components across different subbands. Their analysis derived diagnostic features from multiscale wavelet energies and eigenvalues of subband matrices' covariate analysis for MI classification and localization. Extending this wavelet-based approach, Padhy el al. [39] constructed a third-order tensor structure to represent the multi-lead ECG data and employed higher-order singular value decomposition on the wavelet transformed representation. This approach leveraged intra-beat, inter-beat, and inter-lead correlations, extracting mode-n singular values (where mode-n refers to the singular values obtained from unfolding the tensor along each of its three dimensions - temporal, beat-wise, and lead-wise) and normalized multiscale wavelet energies across subbands for enhanced MI detection and localization. Zeng et al. [46] adopted VMD for signal decomposition and extracted features from phase space reconstruction of Shannon energy envelopes, derived from the initial four intrinsic modes for arrhythmia classification. Despite their extensive application, traditional frequency-based approaches and manual feature engineering methods face several limitations. These techniques often require domain expertise for feature selection and may not capture subtle, discriminative patterns in ECG signals. The handcrafted features might not generalize well across diverse patient populations and varying recording conditions. Moreover, the complex, non-linear nature of cardiac electrical patterns makes it challenging to design comprehensive feature sets that capture all relevant diagnostic information.

1.3.2 Deep Learning in Cardiovascular Diagnosis

Deep learning methods have emerged as a powerful alternative to conventional feature-engineering based approaches. Deep neural networks, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have emerged as powerful tools for ECG analysis, offering the ability to automatically learn hierarchical representations from raw ECG signals [54]. Early work by Kiranyaz et al. [55] demonstrated the potential of CNNs in detecting a wide range of arrhythmias from single-lead ECG recordings. The success of CNNs can be attributed to their ability to capture local patterns and

hierarchical features in ECG signals through their convolutional operations. The architecture of CNNs consists of convolutional layers that apply learnable filters to extract relevant features, pooling layers for downsampling and translation invariance, and fully connected layers for classification [54]. Both 1D and 2D CNN approaches have been explored extensively. 1D CNNs process raw ECG signals directly, with architectures like ResNet and inception-based networks showing promising results [4]. In contrast, 2D CNN approaches first transform the ECG signal into image-like representations, such as spectrograms or multi-lead heat maps, before applying standard computer vision architectures [56, 57]. Pourbabaee et al. [58] employed a CNN-based architecture focusing on screening patients with paroxysmal atrial fibrillation (PAF).

RNNs, on the other hand, are characterized by their recurrent connections, which enable information to persist across time steps, enabling them to capture long-term dependencies in ECG signals [59]. Among the various RNN variants, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks have gained significant attention in the context of ECG analysis. LSTMs represent an extension of traditional RNNs that incorporates memory cells and gating mechanisms. These additional components enable the selective storage and forgetting of information over extended periods, thereby addressing the vanishing gradient problem that often hinders the performance of standard RNNs. The LSTM architecture consists of three main gates: the input gate, forget gate, and output gate. These gates regulate the flow of information into and out of the memory cell, allowing the network to learn and retain relevant information over long sequences [60]. GRUs, on the other hand, provide a simplified alternative to LSTMs while still maintaining the ability to capture long-term dependencies. GRUs combine the forget and input gates into a single update gate and merge the cell state and hidden state [61]. This streamlined architecture reduces the number of parameters and computational complexity compared to LSTMs, making GRUs more efficient and easier to train. By leveraging the temporal context and learning complex patterns over time, RNNs and its variants have demonstrated promising results in a range of ECG analysis tasks, including arrhythmia classification [41], atrial fibrillation detection [62], MI classification [42].

Hybrid approaches [63–65] incorporating both CNN and RNN architectures have also been explored to capture both spatial and temporal dependencies in ECG signals. Among these methods, Liu et al. [66] devised a hybrid framework called MFB-CRNN employing CNN to learn lead specific features from each ECG lead and Bidirectional Long Short-Term Memory (BLSTM) to summarise information from the 12-lead feature representation for MI detection using 12-lead ECG. Han et al. [67] employed CNN-based residual blocks to learn local characteristics and spatial information for each of the 12 leads and then fused the representational features to detect and localize MI. In a subsequent study [68], they expanded

this architecture by incorporating squeeze and excitation networks and a BLSTM network to summarize information from the 12-lead feature representation based on importance, further enhancing MI detection and localization. In a different approach focusing on efficient architecture design, Dissanayake et al [69] introduced DConv-LSTM, a hybrid architecture that combines dilated convolutions with LSTM networks. Their model leverages dilated convolutions in residual blocks to capture multi-scale temporal patterns while maintaining computational efficiency.

More recently, attention-based models have been employed in ECG analysis, such approaches identify informative regions and leverage them to achieve optimal performance. Yu et al. [43] transformed the multilead ECG signal into a 2-D matrix using width-concatenation. They utilized CNN-based dense blocks with self-attention mechanism to capture spatial and temporal features for end-to-end detection of MI from ECG signals. Jyotishi et al. [44] similarly sought to acquire spatio-temporal ECG features by applying multihead criss-cross attention to a feature representation based on LSTM. Yang et al. [70] leveraged multi-scale CNN blocks with coordinate attention modules for multi-view feature representation to enhance multi-label ECG classification. The integration of attention mechanisms in ECG analysis creates dynamic, learnable relationships between different components of the signal. This capability is particularly valuable for capturing both local beat-to-beat variations and global rhythm patterns simultaneously. The attention modules achieve this by computing relevance scores for different signal segments, effectively weighting their contributions to the final diagnostic decision. This dynamic weighting mechanism allows the model to adapt its focus based on the specific characteristics of each ECG recording and the diagnostic task at hand. Furthermore, the attention weights provide valuable interpretability, allowing clinicians to understand which parts of the signal contributed most to the diagnostic decision.

In parallel with attention-based approaches, Multi-task Learning (MTL) has emerged as another powerful paradigm in automated ECG analysis. MTL is a strategy of jointly learning different correlated tasks by optimizing multiple loss functions simultaneously. In the context of ECG analysis, MTL frameworks often encompass a primary diagnostic classification task alongside complementary tasks such as signal denoising, waveform segmentation, and quality assessment. By jointly optimizing these tasks, MTL enables the deep learning model to capture a more comprehensive understanding of the ECG signal, leading to improved generalization and robustness. Lv et al. [71] proposed an MTGBi-LSTM framework for multi-lead ECG signals, employing MTL to enhance diagnostic accuracy for various cardiac pathologies. The tasks included broad disease class and coarser class classification, addressing data imbalances through dynamic loss weighting. Prabhakararao et al. [72] developed a multi-task deep learning framework (MT-DCNN) for estimating the AF burden long-term ambulatory ECG with reconstruction of ECG sequence as an auxiliary

task.

While previous research in automated CVD diagnosis has predominantly relied on digital ECG signals, paper ECG plots remain a major source of clinical data. This prevalence of paper-based ECGs in clinical settings has motivated the development of deep learning methods specifically designed for ECG image interpretation. Among these methods, Hao et al. [73] trained a multi-branch network of shallow CNNs to extract lead-wise features from ECG images and used densenet to perform MI classification. Makimoto et al. [74] adopted a 6-layered CNN on images of 12-lead ECG for myocardial infarction detection and compared its performance with the consensus of 10 cardiologists. These approaches have essentially concentrated on a specific abnormality, framing the task as binary classification. Abubaker et al. [75] employed CNN based backbone networks to extract high level features from ECG images for cardiac abnormality prediction. Gliner et al. [76] proposed a two-way approach of automated cardiac rhythm identification using either signal or image. Two different network architectures for signal and image-based classification were developed to classify multiple cardiac rhythms. More recently, Sangha et al. [77] built a multi-label CNN network based on efficientNet architecture to diagnose six rhythm and conduction disorders from ECG images. They also compared the image-based model with a custom signal-based CNN model and found the performance of the image-based model superior to the signal-based model. Notably, these advanced methods have devised strategies for ECG image analysis tailored to either rhythmic or morphological abnormalities.

1.4 Motivation for the Research Work

Although significant advances have been made in automated ECG-based cardiac diagnosis using deep learning approaches, several critical challenges remain unaddressed in developing reliable and clinically applicable solutions. The key research challenges that motivate this work are as follows:

- Prevalent approaches treat ECG analysis purely as a data-driven task, relying on abstract feature representations acquired at deeper network layers to classify pathology, much like general object classification tasks. The effectiveness of such approaches is constrained by training data quality, and they often fail to incorporate the vast domain knowledge accumulated through decades of clinical practice. These abstract representations, while powerful for pattern recognition, may not effectively capture the subtle temporal relationships and morphological changes in ECG waveforms that are crucial for accurate diagnosis. For instance, in MI detection, clinicians systematically analyze specific ECG components: the morphology of the QRS complex for identifying Q-waves, the elevation or

depression of ST segments for determining acute ischemia, and changes in T-wave patterns for assessing repolarization abnormalities. These components and their inter-relationships provide critical information about the location of infarction, extent of myocardial damage, and disease progression, nuanced clinical insights that may be overlooked by purely data-driven approaches focusing on holistic pattern matching rather than physiologically meaningful features. Therefore, we hypothesize that incorporating such domain knowledge and clinically relevant features into the deep learning framework will result in improved diagnostic performance compared to models trained in a purely data-driven manner, particularly in identifying subtle yet clinically significant patterns that characterize different cardiac conditions.

- As discussed before, cardiovascular diseases often co-occur, yet most existing approaches treat ECG diagnosis as a simple multi-class classification problem, where each ECG recording is assigned to a single diagnostic class. The effectiveness of such approaches is limited by their inability to capture the intricate dependencies and co-occurrence patterns between different cardiac conditions. For instance, patients might simultaneously present with both rhythm and conduction abnormalities, where specific arrhythmias like atrial fibrillation may co-exist with various degrees of AV blocks. Also, In clinical practice, cardiologists perform a comprehensive analysis considering multiple levels of diagnosis, from broad disease categories to specific subtypes, along with potential comorbidities. However, most available methods in the literature treat cardiac diagnosis as a flat classification problem, overlooking the natural hierarchy between general cardiac disorders and their specific subtypes. These hierarchical relationships and disease co-occurrences provide crucial diagnostic information that may be overlooked by approaches that treat ECG classification as a flat multi-class classification problem. Therefore, a more sophisticated approach that can handle multi-label classification and hierarchical disease relationships is needed for comprehensive cardiac diagnosis.

- Existing approaches in automated ECG analysis focus solely on providing deterministic predictions without quantifying the associated uncertainty in their decisions. This limitation becomes particularly critical in clinical settings, where misdiagnosis can have serious consequences for patient care. In clinical practice, cardiologists often assign varying levels of confidence to their interpretations based on ECG signal quality, presence of noise, clarity of diagnostic patterns, and their own expertise. For instance, when analyzing complex arrhythmias or subtle ST-segment changes, clinicians might express different levels of certainty depending on the clarity and consistency of the observed patterns. However, existing deep learning methods typically output point predictions without providing

confidence estimates or reliably identifying cases where they might be uncertain. This becomes particularly problematic when models encounter ECG patterns that deviate from their training distribution or when dealing with borderline cases where multiple interpretations might be plausible. Therefore, we hypothesize that incorporating principled uncertainty quantification methods into deep learning models will lead to more reliable and clinically applicable systems.

- A fundamental challenge in deploying deep learning models for clinical ECG analysis lies in their ability to handle data distributions that differ from their training sets. While current approaches show promising results on standard datasets, they often fail when encountering out-of-distribution (OOD) samples, ECG patterns that deviate significantly from their training distribution. In clinical practice, such variations are common and arise from multiple sources: differences in patient demographics, recording conditions, device specifications, and most critically, encountering cardiac conditions that were not part of the training set. For instance, a model trained to detect specific arrhythmias might encounter a novel cardiac condition during deployment and incorrectly force a prediction into one of its known categories rather than indicating uncertainty. However, existing deep learning methods typically lack mechanisms to identify when they are operating outside their reliable decision boundaries, potentially leading to overconfident but incorrect predictions on OOD samples. This limitation becomes particularly critical in healthcare applications where undetected erroneous predictions could influence clinical decisions. Therefore, we hypothesize that incorporating robust OOD detection mechanisms into deep learning models will enhance their clinical reliability, enabling systems to not only make accurate predictions but also reliably identify cases requiring additional clinical review.

- While significant advances have been made in automated ECG-based cardiac diagnosis using deep learning approaches, these developments have predominantly focused on digital ECG signals. Although these methods have shown promising results, they cannot be directly applied to ECG paper plots, which remain the primary format of ECG records in many clinical settings, particularly in resource-constrained healthcare facilities and for historical patient data. Current approaches to ECG paper plot analysis often treat the task merely as a standard computer vision problem, overlooking critical aspects of clinical ECG interpretation. While paper ECGs present multi-lead information in a structured layout that clinicians leverage for comprehensive diagnosis, existing approaches often fail to effectively utilize these spatial relationships and specialized regions. For instance, clinicians analyze the rhythm strip at the bottom of the ECG plot specifically for rhythm abnormalities, while simultaneously examining the standard 12-lead layout above for morphological

changes and disease patterns. This structured organization, where different regions of the ECG plot serve distinct diagnostic purposes, carries inherent clinical significance that current image-based methods typically overlook.

1.5 Major Contributions of the work

Motivated by these challenges in automated ECG analysis, this dissertation proposes novel deep learning frameworks to address these critical gaps. The key contributions of this work are as follows:

- We introduce the Cross-Task Attention Transfer (CTAT) framework as a novel solution for incorporating clinical domain knowledge into deep learning-based ECG analysis. Specifically designed for robust and context-aware MI detection, the framework leverages a knowledge distillation approach that transfers morphological understanding from ECG delineation to enhance MI detection. Our architecture consists of three principal components: a teacher network based on U-Net with attention and residual connections that captures morphological features through ECG delineation, a student network based on ResNet with spatial attention optimized for MI classification, and a CTAT module that enables knowledge transfer between these networks. Through this architecture, we enable the MI detection model to focus on clinically relevant ECG regions, guided by the morphological knowledge captured in the delineation task. Comprehensive evaluation on three distinct 12-lead ECG databases demonstrates the framework's superiority over existing methods, establishing a new paradigm for integrating clinical knowledge into automated ECG analysis.

- Extending upon our approach of incorporating clinical domain knowledge into deep learning frameworks, we present the Multi-task Physics-Constrained Hierarchical MI Classification framework (MTPCHi-MI), a novel end-to-end solution for improved MI detection and localization. The framework uniquely combines multi-task learning with physics-guided constraints and hierarchical classification to capture the complex nature of ECG diagnosis. Our architecture employs auxiliary tasks of ECG delineation and physics-informed ECG reconstruction alongside the primary task of hierarchical MI classification. The framework consists of three key components: a shared encoder that learns common latent representations across tasks, twin parallel decoders based on U-Net with attention and residual connections for the auxiliary tasks, and an LSTM-based decoder for hierarchical MI classification. A physics-informed ECG reconstruction loss ensures the learned representations preserve the underlying characteristics of 12-lead ECG signals. The autoregressive LSTM decoder enables sequential predictions following the natural hierarchy of MI diagnosis, while the hierarchical

loss function guides the model to respect the inherent relationship between MI detection and its localization, mirroring the clinical diagnostic process. Comprehensive evaluation on three distinct 12-lead ECG databases demonstrates the framework's superior performance over existing methods, establishing a new approach for physics-aware and hierarchically-structured cardiac diagnosis.

- An Uncertainity-Aware ECG diagnostic approach has been investigated for confident multi-label cardiovascular abnormality detection from 12-lead ECG signals. Our approach leverages variational inference to simultaneously achieve accurate classification, uncertainty quantification, and out-of-distribution detection. To address the multi-label nature of cardiac diagnosis, we incorporate a learnable condition interaction matrix that captures interdependencies between cardiac abnormalities, enabling more coherent multi-label predictions. The framework employs a variational autoencoder with Gaussian mixture latent representation that learns the underlying ECG signal distribution. We estimate input signal uncertainty by analyzing the entropy of the posterior distribution over GMM components in the latent space, providing a principled approach to uncertainty quantification. Extensive experiments on external databases demonstrate the framework's ability to reliably detect unfamiliar ECG patterns, providing an essential safety mechanism for clinical deployment.

- Inspired by the clinical interpretation of ECG paper records, where different regions serve distinct diagnostic purposes, the rhythm strip for rhythm analysis and the 12-lead layout for morphological assessment, we present a novel multi-task learning framework for automated ECG image analysis. Our framework simultaneously addresses morphological cardiac ailment classification and cardiac rhythm classification, while incorporating an auxiliary task of region delineation to leverage this structured layout effectively. The architecture consists of three key components : a ResNet-101 backbone network that extracts shared features from the ECG image, a region-based pooling mechanism that separates these features into distinct representations corresponding to the rhythm strip and 12-lead sections, and task-specific classification networks for final diagnosis. We incorporate homoscedastic uncertainty-based dynamic weighting to optimally balance different tasks, while task-specific regions of interest ensure focused feature extraction for each diagnostic aspect. The morphological classification branch employs a multi-label approach to account for potential cardiac comorbidities, reflecting real-world clinical scenarios. Comprehensive evaluation on ECG images derived from an open-access dataset and an internal clinical ECG image database demonstrates our framework's superior performance over single-task approaches and existing state-of-the-art methods, establishing an effective approach for automated interpretation of ECG paper records.

1.6 Organization of the Thesis

The thesis is divided into four working chapters, each focusing on a specific aspect of improving the diagnosis of cardiovascular abnormalities using ECG signals. **Chapter 2** introduces a morphology-aware ECG diagnostic framework that utilizes knowledge distillation to transfer attention from the ECG delineation task to the MI detection task. By incorporating domain knowledge about the important regions in the ECG signal, the framework aims to enhance the accuracy of MI diagnosis. **Chapter 3** builds upon the MI detection task by extending it to MI localization using a hierarchical decision network. The chapter proposes an end-to-end approach that integrates domain knowledge about the important ECG regions into the classification network through a multi-task learning framework. The auxiliary tasks of ECG reconstruction and delineation are employed to facilitate this integration. **Chapter 4** addresses the challenge of multi-label ECG diagnostics by developing an uncertainty-aware framework. The proposed approach utilizes a variational autoencoder with a Gaussian mixture model latent distribution, where components correspond to cardiac conditions, enabling learning of underlying condition relationships. A learnable condition interaction matrix captures interdependencies between cardiac abnormalities for coherent multi-label predictions. Out-of-distribution detection is achieved through reconstruction error analysis and entropy assessment of the posterior distribution over GMM components. **Chapter 5** explores the potential of using ECG paper plot images, which are readily available in hospital settings, for cardiovascular abnormality detection. A multi-task network is proposed to simultaneously perform morphological disease diagnostics and ECG rhythm classification. Region-based delineation is employed to separate regions corresponding to morphology and rhythm in the standard representation of the ECG paper plot. The delineation information is then used to extract task-specific features using rotational ROI align. Homoscedastic uncertainty is utilized to balance the losses between the two tasks. Finally, **Chapter 6** concludes the research work and highlights some future directions.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

2

Morphology-Aware MI Diagnosis

Contents

As introduced in previous chapter, Myocardial Infarction (MI) is a complex and multifactorial disorder that occurs when unstable atherosclerotic plaques in the coronary arteries rupture, leading to blood clot formation and obstruction of blood flow to the heart muscle [17]. In clinical practice, the 12-lead ECG serves as a primary screening tool for MI [78]. The alteration in ECG morphology during MI dynamically evolves over time. The early manifestation of MI on ECG morphology involves the emergence of hyperacute T waves, usually noticeable within minutes of the onset of coronary occlusion. Following the hyperacute T wave, ST-segment elevation typically occurs, usually within 30 minutes of the onset of coronary occlusion [22]. ST-segment elevation is most pronounced in the leads that correspond to the area of the myocardium that is being infarcted. Over time, the ST-segment elevation will resolve, and T-wave inversion will develop. The Q-waves will typically appear within 12 hours of the onset of MI, and they may persist indefinitely [23–25]. Based on these observations and the corresponding leads reflecting these alterations, clinicians can ascertain the location and severity of the infarct region and subsequently plan treatment.

Existing approaches usually follow a data-driven approach [42–44, 79], and do not explicitly take any clinical priors into consideration for detecting MI from ECG signals. Ideally, the classification networks should mimic the experts when predicting the labels for a given ECG. While conventional deep learning models rely on abstract feature representations from deeper network layers, similar to general object classification tasks, this approach may overlook critical pathological patterns specific to MI diagnosis. In clinical practice, cardiologists focus on specific abnormalities in the QRS complex, ST segment, and T-waves to determine infarction location, assess myocardial damage extent, predict outcomes, and guide therapeutic decisions [80, 81]. However, such domain knowledge has not been explicitly considered yet by current deep learning-based methods.

In this chapter, we propose a novel morphology-aware framework that incorporates clinical knowledge into the deep feature learning process, directing the model's attention to critical morphological regions for enhanced MI diagnosis. Our approach is built on the hypothesis that integrating relevant domain knowledge improves model performance compared to purely data-driven methods. We implement this through an attention-based ECG delineation network that captures morphological characteristics of ECG signals, coupled with a cross-task attention transfer (CTAT) framework. This CTAT mechanism distills morphological insights from the ECG delineation task into the MI classification process, effectively embedding clinical knowledge into the diagnostic pipeline. The objective is to make the student network attend more to input regions sensitive to morphological alterations due to MI. Our proposed method introduces a novel approach to transfer this region-specific attention information from teacher network to a student network, enhancing its ability to focus on relevant regions for improved detection of MI. The rest of the chapter is organised

as follows. Section 2.1 discusses the proposed CTAT framework. Section 2.2 presents the experimental setup. Section 2.3 presents the results and discussion of the proposed framework. Section 2.4 provides a summary of the chapter.

2.1 Proposed Method: Cross-Task Attention Transfer for Morphology-Aware MI Diagnosis

The CTAT framework is inspired by the observation that cardiologists typically adhere to specific ECG morphologies when confirming a diagnosis of MI. To effectively utilize morphological insights, we reframe MI classification within a knowledge-distillation framework comprising two primary tasks: ECG delineation and MI detection. Our goal is to enhance MI detection performance by transferring spatial attention patterns from the delineation task. As illustrated in Fig. 2.1, our architecture consists of three key components: a teacher network with spatial attention blocks trained for ECG delineation; a student network focused on MI classification; and a cross-task attention transfer module that guides the student to adopt the teacher's attention patterns. The student network incorporates an attention mechanism similar to the teacher's, as shown in Fig. 2.1(b). The attention transfer process (detailed in section 2.1.4) is designed to minimize differences between the attention maps of both networks, effectively redirecting the student's focus toward clinically relevant ECG regions. These attention maps represent the relative importance assigned to different regions of the input data. Let $Att_T = \{att_T^l\}_{l=1}^L$ be a set of the attention maps generated by the teacher network and $Att_S = \{att_S^l\}_{l=1}^L$ be a set of the attention maps from the student network. Here, L denotes the number of attention blocks, with each block having its own attention map size and channel dimension. The objective function (Equation 2.5) minimizes the divergence between attention distributions of both models, encouraging the student model to adopt the teacher model's attention patterns. By integrating spatial attention learned from ECG fiducial point segmentation into the MI diagnosis network, we provide crucial morphological context that enhances diagnostic accuracy.

2.1.1 Problem definition

Our objective is to enhance ECG-based MI detection by incorporating clinical knowledge through morphological attention. We begin by training a teacher model f^T for ECG delineation using dataset $\mathcal{D}^t$=$\{(x_n^t, y_n^t)|1 \leq n \leq N^t\}$ with N^t instances. Here, x_n^t represents an input instance, and $y_n^t \in \mathcal{R}^{w \times d}$ denotes its corresponding delineation label, where d is the ECG temporal dimension and $w = 3$ represents the number of ECG wave categories (P wave, QRS complex, T wave). For the student network f^s, we use dataset $\mathcal{D}^s$=$\{(x_n^s, y_n^s)|1 \leq n \leq N^s\}$, where $x_n^s \in \mathbb{R}^{12 \times d}$ is a 12-lead ECG input instance with binary label

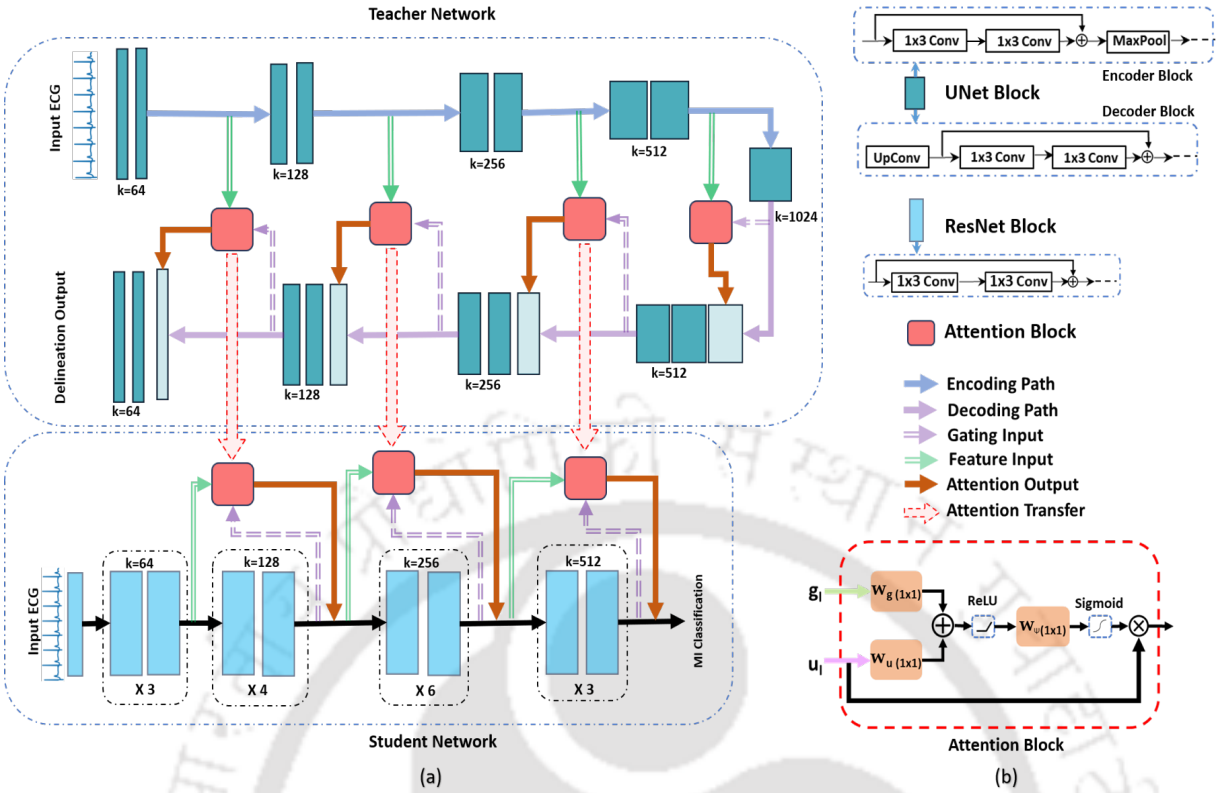

Figure 2.1: (a) Schematic overview of the proposed approach. The teacher network employs a Unet-based architecture to capture morphology-specific attention, while the student network is based on Resnet with attention mechanism for morphology-aware detection of MI. (b) Structure of the attention block.

$y_n^s \in \{0, 1\}$. Our goal is to develop a binary classifier $f^s(x^s) : \mathcal{R}^d \rightarrow \{0, 1\}$ based on $\mathcal{D}^s$ that maps ECG signals to MI diagnosis. The teacher network acquires valuable knowledge about ECG wave morphology through delineation. We then transfer this learned attention, denoted as $Att = \{att^l\}_{l=1}^{L}$, from the teacher to guide the student model's learning process for MI classification.

2.1.2 Structure of the Teacher Network

For the ECG delineation task, the teacher network is based on a residual UNet with spatial attention, a 1D adaptation of the UNet architecture [82] enhanced with residual connections and attention mechanisms. The network maintains the standard UNet structure with an encoding path, decoding path, and skip connections linking corresponding encoder-decoder layers. The encoding path consists of sequential convolutional blocks, each containing two 1×3 convolution layers with ReLU activation, followed by max pooling with kernel size 2 for downsampling. As illustrated in Figure 2.1(a), the network begins with the input ECG signal and proceeds through four downsampling blocks. The number of feature channels progressively increases from 64 in the first block to 128, then 256, and finally 512 in the fourth block, while

the spatial dimensions are correspondingly reduced at each max pooling operation. This architecture enables the network to capture hierarchical features at multiple scales, with shallow layers learning low-level waveform patterns and deeper layers extracting high-level semantic representations. Identity mappings [83] are applied to each block through skip connections, where the block input is directly added to its output, preserving high-resolution feature propagation to deeper layers and mitigating vanishing gradient problems during backpropagation. This hierarchical structure effectively captures multi-scale ECG morphological features.

The decoding path mirrors the encoder's architecture while incorporating skip connections to reconstruct spatial information by upsampling the encoder-generated feature maps. These skip connections link encoder outputs directly to corresponding decoder inputs, providing prior high-resolution information for the upsampling process and aiding accurate reconstruction from the compressed feature representation. Both encoder and decoder blocks utilize residual connections, with a 1×1 convolutional layer (stride=1) ensuring dimensional compatibility between residual outputs and feature maps. At each decoding stage, an attention block (detailed in Section 2.1.4) is integrated to selectively emphasize relevant features during upsampling. The attention block receives two inputs consisting of the decoder output, which serves as a gating signal providing contextual information from coarser scales, and the corresponding encoder features from the skip connection, which contain high-resolution spatial details. Through convolutional layers and non-linear activations, it computes spatial attention weights that indicate the relative importance of different regions. These attention weights are then applied to the encoder features through element-wise multiplication, enabling the model to selectively emphasize morphologically relevant ECG regions. The complete teacher network incorporates four attention blocks, enhancing its ability to precisely delineate ECG fiducial points. We optimize this architecture using binary cross entropy loss (Equation 2.1).

$$\mathcal{L}_T = -\frac{1}{M}\sum_{w=1}^{W}\sum_{m=1}^{M}(y_{n,w}^{t}\log P_t) + (1-y_{n,w}^{t})\log(1-P_t)) \tag{2.1}$$

where P_t is the prediction probability score, M is the total number of mini-batch samples and w represents the number of waves.

2.1.3 Structure of the Student Network

For the MI classification task, we implement a 1D adaptation of ResNet34 with attention mechanisms. ResNet34 was specifically selected to ensure dimensional compatibility with the teacher network's U-Net architecture, enabling direct attention transfer at three matching scales (64, 128, and 256 channels)

without requiring additional alignment operations. The backbone architecture consists of 16 sequential residual blocks with identity mappings, where each block contains two convolutional layers with 1×3 kernel dimensions. The structure is organized into four distinct stages, each characterized by a specific number of residual blocks and unique output channel dimensions. This backbone network effectively transforms inputs from the original ECG signal space $\mathcal{X} \in \mathbb{R}^{12\times d}$ into a latent feature space $\mathcal{Z}$. The output representation from the l^{th} residual block is formulated as

$$\mathcal{Z}_l = F(I_{l-1}, \{\Theta_l\}) + I_{l-1} \tag{2.2}$$

where $F(I_{l-1}, \{\Theta_l\})$ is the residual with parameters Θ_l. Essentially, F computes a residual rather than the output I_l directly as the shortcut connection from the input I_{l-1} is added to the output I_l of the residual block. This bypasses the convolution layers in the module and performs identity mapping with the module output. Fig. 2.2 illustrates this structure with explicit notation. Attention Module is added to each of the convolutional stages other than the first stage to avoid low-level feature maps and match the attention structure of the teacher network.

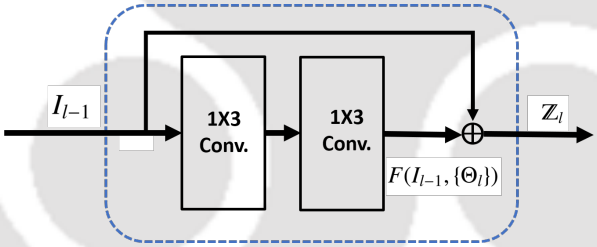

Figure 2.2: Residual block structure showing input features I_{l-1}, two 1×3 convolutional layers forming the residual function $F(I_{l-1}, \{\Theta_l\})$, and output features $\mathcal{Z}_l$ obtained through element-wise addition ($\oplus$) with the identity mapping.

2.1.4 Cross-Task Attention-Transfer Module

The proposed CTAT module enhances essential fiducial region information while suppressing non-fiducial features across multiple scales. The attention mechanism utilized in this work operates on a principle similar to the one employed in [84], is illustrated in Fig. 2.1(b) and is shared by both teacher and student networks. Let $u^l \in \mathbb{R}^{C_u\times H\times W}$ be the feature maps from the l^{th} encoder block and $g \in \mathbb{R}^{C_g\times H\times W}$ be the gating input, where H and W represent spatial dimensions while C_u and C_g indicate channel dimensions of the input features and gating signal respectively. In the teacher network, the gating signal g comes from deeper layers in the decoding path, which have undergone multiple downsampling operations in the encoder and capture coarse-grained, high-level semantic representations of the entire ECG signal.

These deeper features encode global patterns such as overall heart rhythm, the presence and approximate locations of different waveforms, and their relative temporal relationships. Similarly, in the student network, the gating signal comes from the output of the same convolutional stage, which has processed and aggregated information across larger receptive fields through successive convolutions, providing contextual understanding about broader temporal patterns in the ECG signal. By combining this contextual information (g) with the local feature representation (u^l) through the attention mechanism, the network can make more informed decisions. The attention block computes coefficients α^l that selectively emphasize spatial regions in u^l across all feature channels based on the broader context captured in g. These attention coefficients α^l are computed as follows:

$$q_{att}^l = W_\psi^T(ReLU(W_u^T u^l + W_g^T g + b_g))) + b_\psi \tag{2.3}$$

$$\alpha^l = \sigma(q_{att}^l(x^l, g; \Theta_{att})) \tag{2.4}$$

where σ corresponds to the sigmoid activation and $W_\psi^T \in \mathbb{R}^{1\times C_{int}}, W_u^T \in \mathbb{R}^{C_{int}\times C_u}, W_g^T \in \mathbb{R}^{C_{int}\times C_g}, b_g \in \mathbb{R}^{C_{int}}, b_\psi \in \mathbb{R}$ are the set of parameters (Θ_{att}) for the attention block, which are optimized during training. W_u^T and W_g^T project the input feature representation and the gating input onto an intermediate feature space of dimension C_{int} through 1×1 convolutional operations. These projected features are element-wise summed, and the attention coefficients are obtained by employing the learned weights W_ψ^T and bias b_ψ.

The attention coefficients undergo normalization using the sigmoid activation function, constraining to be in a range of 0 and 1, where higher values correspond to more salient regions in the input features. These coefficients modulate the input representation u^l through element-wise multiplication $\hat{u}^l = u^l \cdot \alpha^l$ to produce the attention block output. The attention output is then added to the section's output before proceeding to the next sequential section. Fig. 2.3 illustrates the detailed architecture of each attention block. To effectively transfer attention knowledge from teacher to student network, we aim to align the activations between corresponding attention blocks in both networks. We employ Kullback-Leibler (KL) divergence [85] to quantify and minimize the discrepancy between attention distributions across the networks. If we denote att_T^l and att_S^l as the normalized attention maps for layer l in the teacher and student networks respectively, the KL-divergence loss is formulated as:

$$\mathcal{L}_{CTAT} = \frac{1}{M}\sum_{l=1}^{L}\sum_{m=1}^{M} p(att_T^{l,m}) \log \frac{p(att_T^{l,m})}{p(att_S^{l,m})} \tag{2.5}$$

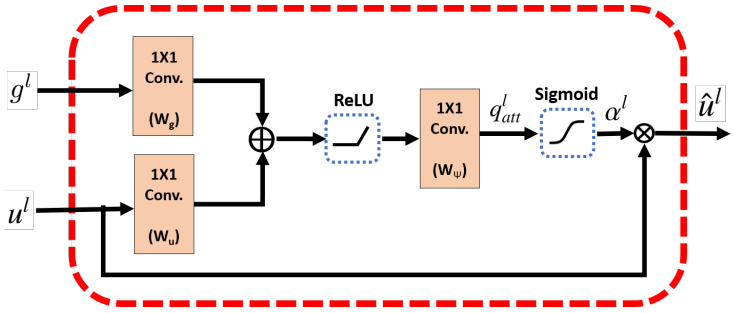

Figure 2.3: The architecture of the attention block. A gating signal (g^l) and an input feature map (u^l) are processed to compute attention coefficients (α^l). These coefficients re-weight the input features to produce the attended output $\hat{u}^l$.

where $p(\cdot)$ represents the softmax function, defined as $p(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$ for each element x_i. This converts the activation values into a probability distribution. As observed from Eq. 2.5, the KL divergence is an asymmetric loss. When $p(att_T)$ is high, there is a need for $p(att_S)$ to be high to minimize the loss. Conversely, for a small $p(att_T)$, the value of $p(att_S)$ has little impact on the overall loss [86, 87]. Consequently, the KL divergence makes the student network's attention to closely match that of the teacher network, particularly for high activations. This emphasizes the prominent features while dampening irrelevant ones.

The training procedure for our proposed CTAT framework begins with pre-training the teacher model on the ECG delineation task using the loss function $\mathcal{L}_T$ defined in Equation (2.1). Following this initial phase, we freeze the teacher network parameters and proceed to train the student network, optimizing only its parameters. During this stage, we utilize the attention maps generated by the teacher network for attention distillation, without directly using the teacher's predictions in the student's training process. The student network attempts to mimic these attention maps while processing identical input data, with these attention patterns serving as guidance for the student's learning process. The complete loss function for training the student network can be expressed as:

$$\mathcal{L}_S = \beta \mathcal{L}_{CE} + (1 - \beta)\mathcal{L}_{CTAT} \tag{2.6}$$

with

$$\mathcal{L}_{CE} = -\frac{1}{M}\sum_{m=1}^{M}(y^d_{n,c} \log P_d) + (1 - y^d_{n,c}) \log(1 - P_d)) \tag{2.7}$$

Here β provides for balancing between the classification loss and the CTAT loss. It essentially governs the impact of attention transfer between teacher and student network. Note that the parameter β ranges

from 0 to 1, where β=1 corresponds to the configuration with no attention transfer.

2.2 Experimental Setup

This section presents the comprehensive experimental setup used to evaluate the CTAT framework. First, we describe the datasets utilized in this study along with the preprocessing steps performed to prepare the data. Following this, we detail the comparative analysis conducted to assess the framework's performance. We then outline the evaluation metrics and methods employed to measure the effectiveness of the proposed approach. Finally, we conclude the section by presenting the implementation details of the framework.

2.2.1 Clinical ECG Datasets

We utilized two distinct sets of datasets: ECG delineation datasets for training the Teacher network, and MI classification datasets for training the Student network. The specific details of these datasets and their characteristics are discussed as follows.

2.2.1.1 ECG delineation Datasets

Lobachevsky University database (LU database): The LU database [88] database comprises 200 twelve-lead ECG recordings collected at Novgorod City Hospital between 2017 and 2018. Each record spans 10 seconds with a sampling rate of 500 Hz. The database features comprehensive expert annotations performed by experienced cardiologists for all recordings, marking the boundaries and peaks of QRS complexes, P waves, and T waves across each lead.

QT database: The QT database [89] contains 105 two-lead ambulatory ECG recordings representing a diverse patient population with various cardiac conditions. The recordings were strategically selected from existing databases (including the MIT-BIH Arrhythmia Database, the European ST-T Database) to include a wide spectrum of QRS and ST-T morphologies, thereby providing comprehensive representation of different pathophysiological conditions. Each recording captures 15 minutes of ECG sampled at 250 Hz. The database features dual annotation schemes: a manual set created by cardiologists who carefully selected and annotated 30-50 representative beats per recording, and an automated set [90] that provides waveform boundary delineations for every beat in the dataset.

The LU-dataset exclusively serves as the training dataset for the teacher network, while the QT dataset is utilized as the testing dataset, facilitating assessment of the teacher network's generalization performance across different datasets.

2.2.1.2 MI Classification Datasets

Physikalisch Technische Bundesanstalt Extra Large (PTB-XL) : PTB-XL [91] is an extensive 12-lead ECG dataset comprising 21,837 recordings from 18,885 patients, with each record capturing 10 seconds of cardiac activity available at both 500 Hz and 100 Hz sampling frequencies. The database encompasses a comprehensive range of cardiovascular diseases with numerous co-occurring pathologies, while also including a good proportion of healthy ECG recordings for baseline comparison. Furthermore, the authors have organized the database into ten non-overlapping folds using stratified sampling to ensure patient-wise separation, preventing data from the same individual from appearing in multiple folds. They recommended a specific partition for analysis and comparison of automated approaches: folds 1–8 for training, fold 9 for validation, and fold 10 for testing. Throughout our experimental work, we adhered to this recommended data split to ensure reproducibility and comparability with other studies.

Shandong Provincial Hospital (SPH) : SPH [92] contains 25,770 twelve-lead ECG recordings collected from 24,666 patients at Shandong Provincial Hospital between 2019 and 2020. The recordings feature a sampling frequency of 500 Hz with variable durations ranging from 10 to 60 seconds. The dataset encompasses a diverse spectrum of CVDs while also maintaining a good proportion of healthy ECG recordings, providing valuable comparative baseline data for clinical ECG research.

China Physiological Signal Challenge (CPSC) : The third data source is the challenge dataset of CPSC 2018, which was held during the 7th International Conference on Biomedical Engineering and Biotechnology in Nanjing, China. This dataset contains ECG recordings collected across 11 hospitals, with recording durations varying from 6 to 144 seconds, all sampled at 500 Hz. For our study, we utilized two specific components: the original public training dataset (CPSC) and a previously unused supplementary dataset (CPSC-Extra). Together, these sources provide 13,256 ECG recordings, including 1,034 cases of MI and 984 healthy control ECGs.

2.2.1.3 Preprocessing

ECG signals from all databases underwent a standardized preprocessing pipeline. First, we removed baseline wander using a moving average filter [93] that computes the average within a defined window as it moves along the signal, creating a smoothed baseline that is then subtracted from the original ECG. Next, we applied a Non-local means (NLM) filter [94] to eliminate muscle artifacts and high-frequency noise. The NLM filter operates on the principle of non-local similarity, computing weighted averages of samples within a search window, where the weight for each sample is determined by the similarity between its

surrounding neighborhood and the neighborhood of the target sample. This approach effectively exploits the repetitive patterns present in ECG signals across heartbeats, preserving morphological features while suppressing noise. The signals were then normalized using Z-score normalization and assessed for quality based on established criteria [95]. The ECG signals in the QT database have a sampling frequency of 250 Hz. To ensure frequency consistency across all datasets, signals from the remaining databases were downsampled to 250 Hz. For the ECG delineation task, we extracted uniform 10-second segments from all records and generated binary sequences corresponding to P-wave, QRS complex, and T-wave annotations to serve as delineation ground truth. For each waveform category, the binary sequence contains values of 1 for sample positions within the annotated waveform boundaries (onset to offset) and 0 elsewhere. During delineation model training, we implemented data augmentation through three techniques: random periodic spikes, additive white Gaussian noise (AWGN), and random flips about the baseline. Random periodic spikes simulate common ECG artifacts, resembling noise from muscle contractions or electrode movements. These spikes are randomly introduced at different time intervals and with varying amplitudes. Mathematically, this can be represented as: $x_{aug}(t) = x(t) + \sum_i A_i \delta(t - t_i)$, where $\delta(t)$ is the Dirac delta function, A_i is the amplitude of the i^{th} spike, and t_i is the random time instance at which the i^{th} spike is introduced. AWGN introduces random fluctuations and variations similar to those that might arise from electronic noise or other sources of interference. Mathematically, AWGN can be represented as $x_{aug}(t) = x(t) + \mathcal{N}(0, \sigma^2)$, where $\mathcal{N}(0, \sigma^2)$ represents Gaussian noise with zero mean and variance σ^2. The noise variance is adaptively scaled relative to the signal's standard deviation, ensuring appropriate noise levels across ECG recordings with varying amplitudes. Random flips about the baseline involve randomly changing the polarity of the ECG signal, mimicking the polarity variations often seen in ECG recordings. This is achieved by randomly changing the signal's polarity, expressed as $x_{\text{aug}}(t) = -x(t)$. For the MI classification task, we similarly extracted non-overlapping 10-second segments from each record. We utilized training folds 1-8 of PTB-XL for training the student network while reserving the SPH and CPSC datasets exclusively for testing, specifically to evaluate the generalization performance of our proposed method across different clinical environments and recording protocols.

2.2.2 Comparative Analysis

A comprehensive series of comparison and ablation experiments are conducted to evaluate the effectiveness of our proposed method. The following sections detail our experimental design and findings.

- **Ablation Study:** Through a methodical ablation study, we demonstrate the effectiveness of attention transfer in our proposed framework. We establish a baseline consisting of the student model

architecture without any attention blocks. We then progressively evaluate performance enhancements by first integrating spatial attention blocks into this baseline independent of any attention transfer, and subsequently examining the specific influence of our CTAT mechanism on the classification network's performance.

- **Sensitivity Analysis:** To investigate the impact of the hyperparameter β (in Eq.5), we evaluate our model by varying the value of β between 0 and 1. β acts as a balancing between the classification loss and the CTAT loss and controls the degree of attention transfer between the student and teacher networks.

- **Qualitative Analysis:** To gain deeper insights into our network's decision-making process and visualize the input regions receiving highest attention, we employed Grad-Cam [96] visualization techniques. This post-hoc interpretation method allows us to generate and analyze attention heatmaps for our network. We specifically compare these visualizations against those from a network with identical architecture and spatial attention mechanisms but lacking the cross-task attention transfer component, enabling direct assessment of how CTAT influences the network's focus on diagnostically relevant ECG regions.

- **Quantitative Comparison:** We evaluated our proposed approach against several established methods for MI detection from multi-lead ECG signals. The comparison includes five deep learning-based state-of-the-art techniques: Spa-Tem MI [43], ASTLNet [44], ML-Net [79], ML-ResNet [67], and MFB-CBRNN [66], as well as two feature engineering-based machine learning approaches: Ms-HoSVD [39] and MEES [38]. To ensure fair comparison, we standardized the input ECG segment length and loss criteria across all methods to match our proposed framework.

2.2.3 Evaluation Methods

For delineation performance assessment, we utilized precision, recall, F1 score, and relative timing error metrics at multiple tolerance thresholds (25, 40, 70, and 150 ms) for both wave onset and offset detection. Tolerance defines the acceptable time window within which a prediction is considered correct. Relative timing error is reported as mean and standard deviation ($\mu \pm \sigma$) of the difference between actual and predicted boundaries, calculated only for instances where both predicted and ground truth labels fall within the specified tolerance window. To evaluate MI classification performance, we employed a comprehensive set of seven metrics: accuracy, precision, recall, F1 score, AUROC [97], PRAUC [98], and negative log likelihood (NLL) [99]. AUROC quantifies class separability by measuring the model's

ability to distinguish between MI and non-MI cases across all possible thresholds. The ROC curve plots True Positive Rate (TPR) against False Positive Rate (FPR), and AUROC is computed as: AUROC $= \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR})$. PRAUC assesses the precision-recall trade-off at varying probability thresholds, which is particularly valuable for imbalanced datasets. The PR curve plots Precision against Recall, and PRAUC is computed as: PRAUC $= \int_0^1 \text{Precision}(\text{Recall})\, d(\text{Recall})$. NLL measures the dissimilarity between predicted probability distributions and true probability distributions. For binary classification, NLL is defined as: NLL $= -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)]$, where N is the number of samples, $y_i \in \{0,1\}$ is the true label for sample i, and p_i is the predicted probability that sample i belongs to the positive class. Lower NLL values indicate predictions with higher confidence in correct classifications.

2.2.4 Implementation Details

The proposed method was implemented in Python 3 using PyTorch (2.0.1) on a workstation with an Intel Xeon 32-core CPU, 64GB RAM and an NVIDIA A100-PCIE GPU, running Ubuntu 20.04 operating system. The CTAT loss is minimized using the SGD optimizer with momentum 0.9, employing an initial learning rate of 0.01 with a mini-batch size of 32. A scheduler was employed to reduce the learning rate by a factor of 5 if no improvement in validation loss was observed for ten consecutive epochs. We employ an early-stopping approach to prevent the model from overfitting, in which the training is stopped if no improvement in validation loss is observed for 25 consecutive epochs. The model with the minimum validation loss during the optimization process is selected for further analysis.

2.3 Evaluation of the proposed CTAT Framework

This section presents a comprehensive analysis of the experimental results to evaluate the effectiveness of the proposed CTAT framework. We begin by examining the ECG delineation performance of our approach, followed by an assessment of its capabilities in MI detection. We then provide qualitative analysis to demonstrate the interpretability of our framework. Further, we conduct ablation studies and sensitivity analysis to validate the contribution of each component. Finally, we present a quantitative comparison of our framework with various state-of-the-art methods to establish its comparative advantages.

2.3.1 Performance on ECG Delineation Task

In this section, we evaluate the segmentation network's performance for ECG delineation. For training, the ECG signals from the LU database are preprocessed using the preprocessing pipeline described in section 2.2.1.3. The signals thus obtained are split into the train and validation sets. The training set, including

augmented variants, comprised 7680 ECG signals, while the validation set contained 1920 signals. Table 2.1 presents waveform delineation performance across the QT database. We conducted evaluations at four error tolerance windows: 25 ms, 40 ms, 70 ms, and 150 ms, where tolerance represents the acceptable margin of error for predictions to be considered accurate. At 150 ms tolerance, the model achieved its best performance on QRS delineation, with onset and offset F1 scores of 95.50% and 95.98%, respectively. For T-wave delineation, the model reached onset and offset F1 scores of 94.59% and 95.72%, followed by P-wave delineation with onset and offset F1 scores of 88.49% and 89.60%. The model maintained consistent performance across the more stringent tolerance values of 25 ms, 40 ms, and 70 ms. Notably, under the most demanding 25 ms tolerance condition, P-wave delineation precision exceeded that of QRS and T waves. Fig.2.4 illustrates multiple examples of delineation predictions by the teacher network, confirming the model's efficacy in accurately identifying P-waves, QRS complexes, and T-waves.

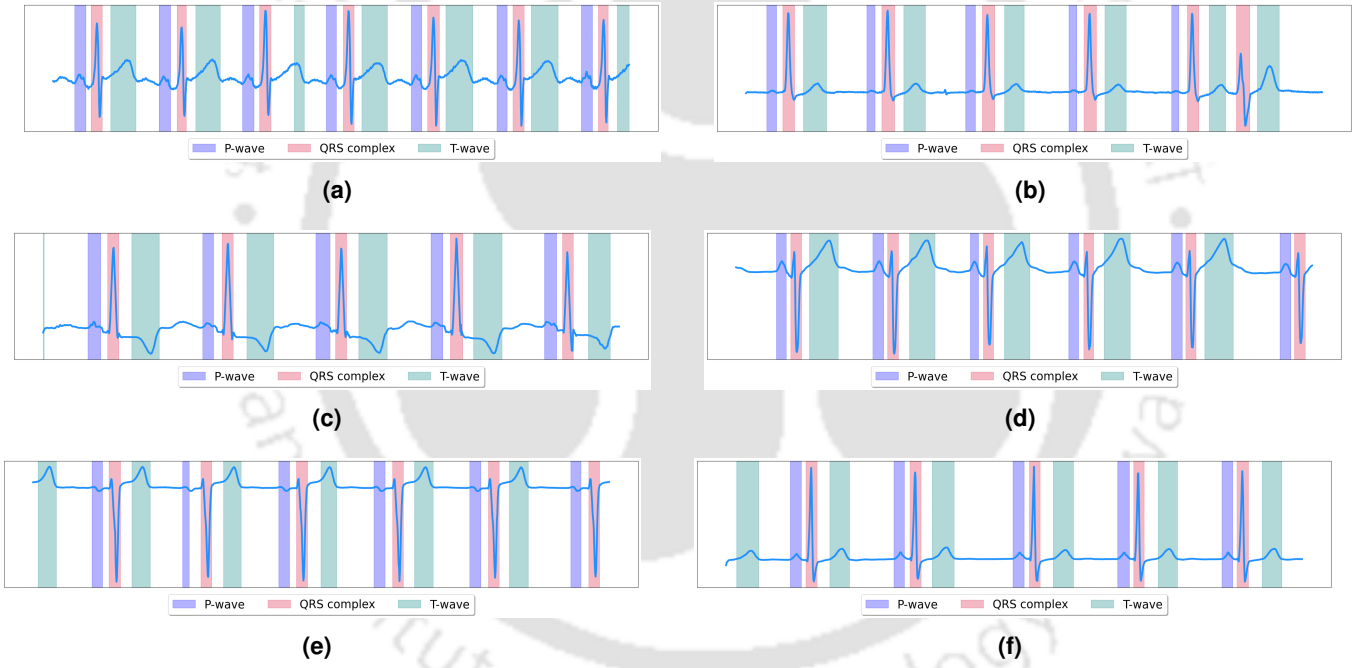

Figure 2.4: Instances of model prediction for P-wave, QRS Complex and T-wave delineation from the QT dataset.

2.3.2 Performance on MI Detection

For MI classification, we trained our student network using preprocessed ECG signals from the PTB-XL dataset, following the recommended data split: folds 1-8 for training (8879 ECG signals; MI: 4379, HC: 4500), fold 9 for validation (1497 signals; MI: 541, HC: 956), and fold 10 for testing (MI: 551, HC: 963). To assess generalization capabilities, we additionally evaluated our model on two external datasets: SPH (MI: 260, HC: 13905) and CPSC (MI: 917, HC: 1503), which were exclusively reserved for testing, with the

Table 2.1: ECG delineation performance on QT database.

		Precision ↑				Recall ↑				F1-Score ↑			
ECG waves	TOL(ms)→	25	40	70	150	25	40	70	150	25	40	70	150
P	Onset	97.86	98.68	98.78	98.83	74.46	78.18	79.50	80.12	84.57	87.25	88.10	88.49
	Offset	98.16	98.84	98.91	98.95	77.30	80.34	81.30	81.87	86.49	88.64	89.25	89.60
QRS	Onset	97.42	98.09	98.12	98.13	92.80	92.83	92.96	93.01	95.06	95.39	95.47	95.50
	Offset	97.56	98.23	98.26	98.27	93.42	93.65	93.75	93.79	95.44	95.89	95.95	95.98
T	Onset	84.19	88.67	92.74	96.26	71.24	80.08	86.78	92.97	77.17	84.16	89.66	94.59
	Offset	95.73	96.45	96.74	97.02	91.42	93.41	93.95	94.46	93.52	94.91	95.33	95.72

		Error($\mu \pm \sigma$) ↓			
ECG waves	TOL(ms)→	25	40	70	150
P	Onset	2.59±1.76	3.20±2.38	4.12±3.67	4.81±5.06
	Offset	2.21±1.91	2.82±2.54	3.58±3.58	4.32±5.23
QRS	Onset	2.10±1.57	2.24±1.80	2.49±2.42	2.62±2.93
	Offset	2.03±1.69	2.27±2.04	2.51±2.53	2.64±3.06
T	Onset	3.34±1.85	4.53±2.60	8.47±5.21	17.45±10.38
	Offset	2.18±1.85	2.75±2.46	3.82±3.94	5.66±7.18

explicit goal of assessing the generalization performance of the proposed framework. Table 2.2 summarizes the quantitative performance of our student network for MI classification. On the PTB-XL test set, the model achieves overall accuracy, F1-score, and AUROC of 92.20 %, 92.17 %, and 97.12 %, respectively with NLL of 0.2063. Notably, performance remained consistent across all three datasets, with particularly strong results on the SPH database where NLL reached 0.1384. This lower NLL value indicates higher confidence in predictions, suggesting the model makes correct predictions with greater certainty. Another notable observation is the consistently high recall across the databases, suggesting a low likelihood of false negatives. This is particularly important in the context of medical diagnosis.

Table 2.2: MI classification performance of the student network on different datasets.

			Predicted (%)							
Database			HC	MI	Accuracy↑	Precision ↑	Recall ↑	F1 Score ↑	AUROC ↑	PRAUC ↑	NLL ↓
PTB-XL (Test-Fold)	Actual	HC	95.01	4.98	92.20	92.17	92.20	92.17	97.17	96.16	0.2063
		MI	12.70	87.29							
SPH	Actual	HC	95.38	4.61	95.41	98.63	95.41	96.62	99.40	87.08	0.1384
		MI	2.69	97.30							
CPSC	Actual	HC	94.22	5.77	92.47	92.77	92.47	92.53	97.96	98.75	0.2011
		MI	8.58	91.41							

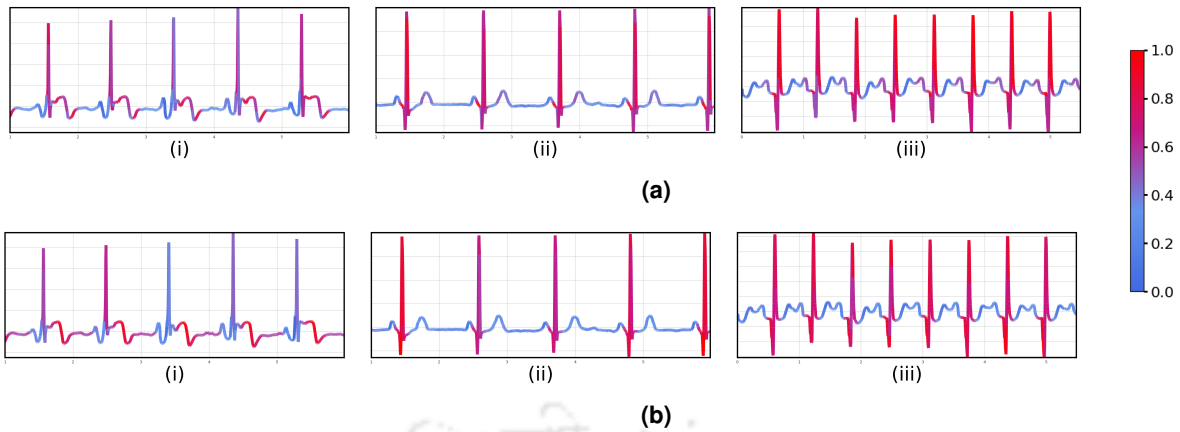

Figure 2.5: Visualization Grad-CAM heatmaps obtained by (a) Resnet with attention and (b) our proposed CTAT network for three subjects with MI diagnosis from PTB-XL Test Set (i,ii,iii) .

In Fig. 2.5, we visualize attention weights to validate our network's ability to emphasize morphologically significant ECG segments, confirming the model's capacity to learn discriminative representations for MI classification. We applied Grad-CAM [96] to generate attention-weight heatmaps for model predictions, analyzing 5-second ECG segments (lead II) from three MI-diagnosed subjects in the PTB-XL test set (Fig. 2.5 (i-iii)). Fig. 2.5(b) demonstrates that our proposed CTAT network directs attention toward ECG regions most sensitive to MI indicators. Specifically, Fig. 2.5(b)(i) shows more weight is given to the the T-wave region, corresponding to visible T-wave inversion in the signal. Similarly, Fig. 2.5(b)(ii-iii) shows focused attention around Q-wave regions, accurately identifying pathological Q-waves characteristic of MI. When comparing heatmaps between the baseline ResNet with attention (baseline+attn.) (Fig. 2.5(a)) and our proposed CTAT network (Fig. 2.5(b)), notable differences emerge in both attention strength and spatial distribution. Our CTAT model exhibits stronger and more concentrated attention on fiducial regions critical for MI diagnosis. Although the attention learned by the ResNet with attention model appears to be similar, it is acquired solely through a data-driven approach without additional task-related context. Such an approach often leads to limited generalizability of the model [100]. The enhanced performance of our CTAT network demonstrates how incorporating morphological attention provides more precise and robust feature focus, directly contributing to improved classification performance.

2.3.3 Ablation Study and Sensitivity Analysis

Table 2.3 summarizes our ablation results. The CTAT framework substantially outperforms the baseline, achieving a 3.04% increase in F1 score, along with improvements of 3.12%, 1.55%, and 3.72% in precision, AUROC, and PRAUC, respectively. This indicates that the addition of morphological information in the form of spatial attention is beneficial for MI diagnosis. To verify that performance improvements specifically

Table 2.3: Ablative study of different components in our proposed method

Configuration			Predicted (%) HC	Predicted (%) MI	Accuracy↑	Precision ↑	Recall ↑	F1 Score ↑	AUROC ↑	PRAUC ↑	NLL ↓
Baseline	Actual	HC	94.91	5.08	89.63	89.65	89.63	89.49	96.41	95.03	0.2685
		MI	19.60	80.39							
Baseline + Attn.	Actual	HC	95.95	4.04	91.87	91.89	91.87	91.79	97.20	96.05	0.2071
		MI	15.24	84.75							
Baseline + Attn . + CTAT (Proposed)	Actual	HC	95.01	4.98	**92.47**	**92.77**	**92.47**	**92.53**	**97.96**	**98.75**	**0.2011**
		MI	12.70	87.29							

result from attention knowledge transfer from the teacher network, we compared our approach with a student network architecture lacking attention transfer (baseline+Attn.). While this intermediate model showed better performance than the baseline, it consistently underperformed compared to our proposed approach across all metrics. This confirms that the CTAT module effectively integrates morphological knowledge from the teacher network into the student network, providing valuable clinical context and enhancing diagnostic performance. We further investigated the optimal balance between attention transfer loss ($\mathcal{L}_{CTAT}$) and classification loss ($\mathcal{L}_{CE}$) in Eq. (2.6) by systematically varying the hyperparameter β from 0 to 1 in increments of 0.1 on the PTB-XL validation set. The analysis revealed that $\beta = 0.7$ provided the optimal balance, achieving the best validation performance while maintaining stable training dynamics. This configuration assigns 70% weight to the classification loss, ensuring strong discriminative capability for MI detection, and 30% weight to the attention transfer loss, providing meaningful morphological guidance without overwhelming the primary classification objective.

2.3.4 Performance Comparison with the Existing Methods

We further validated our approach through comprehensive comparison with five state-of-the-art deep learning methods for ECG-based MI detection: Spa-Tem MI [43], ASTLNet [44], ML-Net [79], ML-ResNet [67], and MFB-CBRNN [66], as well as two conventional feature engineering-based machine learning approaches: MS-HoSVD [39] and MEES [38]. The results are reported on the PTB XL test set, SPH and CPSC dataset to evaluate the generalizability. As shown in Table 2.4, our morphology-aware attention transfer strategy achieves superior performance on the PTB-XL and CPSC datasets and competitive results on the SPH dataset. Attention-based approaches like Spa-Tem MI and ASTLNet, which leverage spatio-temporal ECG representations, demonstrate strong performance across all three datasets. However, their purely data-driven approach slightly limits their effectiveness. In contrast, our CTAT framework leverages

Table 2.4: Comparison Results of our proposed method and other existing state-of-the-art methods for MI classification

Method	PTB-XL (Test fold)				SPH				CPSC			
	F1 Score ↑	Recall ↑	AUROC ↑	NLL ↓	F1 Score ↑	Recall ↑	AUROC ↑	NLL ↓	F1 Score ↑	Recall ↑	AUROC ↑	NLL ↓
Spa-Tem MI [43]	90.31	90.35	96.42	0.2306	91.26	86.44	96.92	0.3282	87.74	87.68	94.42	0.2996
ASTLNet [44]	90.94	91.01	96.67	0.2358	96.61	95.38	98.57	**0.1239**	91.53	91.44	97.87	0.2292
ML-Net [79]	83.51	84.14	91.03	0.3703	94.86	92.64	93.31	0.2451	79.83	79.54	91.26	0.4625
ML–ResNet [67]	89.51	89.63	96.09	0.2902	96.33	94.96	98.80	0.2319	90.03	89.91	97.19	0.2741
MFB-CBRNN [66]	89.73	89.82	95.88	0.2386	94.30	91.57	97.80	0.2249	91.35	91.32	97.20	0.2082
MS-HoSVD [39]	75.49	75.52	83.11	0.5011	86.30	78.14	82.03	0.5417	72.76	72.69	79.30	0.5326
MEES [38]	79.11	79.12	87.32	0.4313	90.16	84.65	89.61	0.4260	76.84	76.50	86.14	0.4947
This Work	**92.17**	**92.20**	**97.17**	**0.2063**	**96.62**	**95.41**	**99.40**	0.1384	**92.53**	**92.47**	**97.96**	**0.2011**

additional clinical context through morphological attention distillation, enabling it to consistently outperform previous state-of-the-art methods. Conventional feature engineering-based methods showed relatively lower performance compared to deep learning approaches. This performance gap can be attributed to deep learning models' ability to automatically extract discriminative feature representations from raw ECG signals, capturing complex patterns and nonlinear dependencies that are difficult to encode using manually engineered features. Fig. 2.6 provides performance visualization through receiver operating characteristic curves (ROC) and precision-recall curves (PR) for all compared methods.

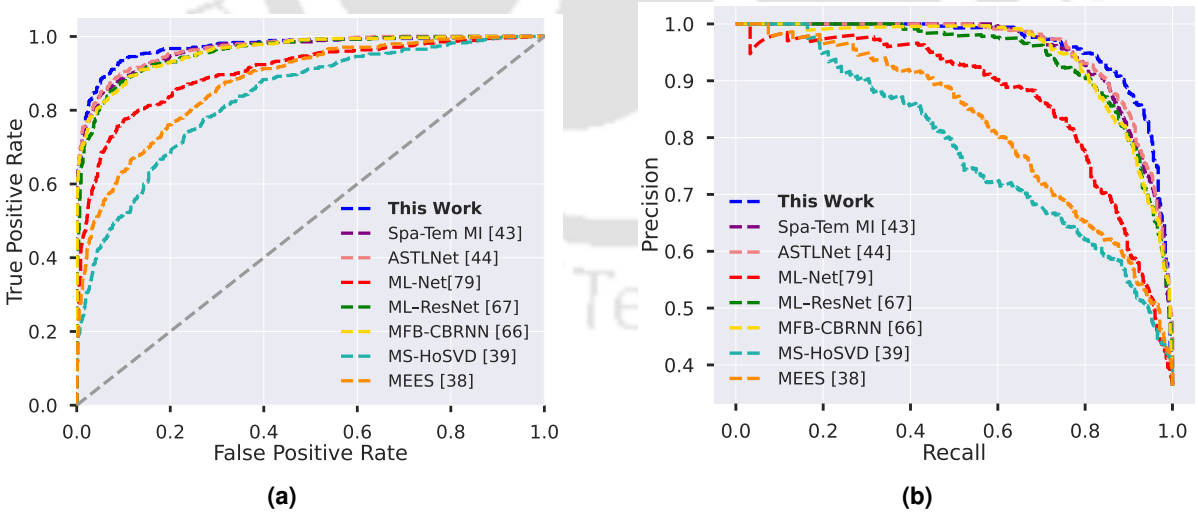

Figure 2.6: (a) The receiver operation characteristic (ROC) curve and (b) precision-recall curve of the proposed CTAT network and other state-of-the-art methods for MI classification.

To evaluate computational efficiency, we measured inference time across state-of-the-art deep learning

methods on an NVIDIA A100 GPU over 1000 iterations (Fig. 2.7). The proposed CTAT framework achieves an inference time of 8.8 ± 1.0 ms per input ECG segment of 10-second, which is competitive with other methods while maintaining superior diagnostic performance. At a sampling frequency of 250 Hz, new 10-second segments arrive every 10 seconds, making our 8.8 ms inference substantially faster than data acquisition and ensuring real-time processing capability. Notably, our method is approximately 1.8 times faster than ML-ResNet [67] (16.2 ± 1.5 ms) while achieving 2.6 % higher F1-score, and maintains competitive latency with Spa-Tem MI [43] (8.7 ± 2.3 ms) and ASTLNet [44] (9.7 ± 0.8 ms) while outperforming both in diagnostic accuracy. Traditional machine learning approaches were excluded from this comparison as they operate on a different computational paradigm, relying on CPU-based handcrafted feature extraction rather than GPU-optimized end-to-end inference, and achieve relatively lower diagnostic performance. The computational efficiency of our CTAT framework, combined with its superior diagnostic performance and clinically interpretable predictions, demonstrates its suitability for deployment in real-time clinical ECG monitoring systems.

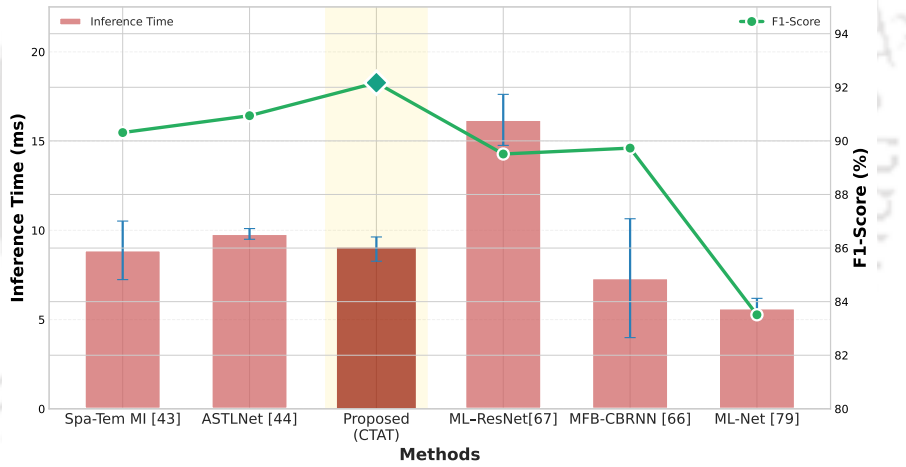

Figure 2.7: Inference time comparison across deep learning methods for MI detection. Bars represent mean inference time with standard deviation (error bars) per 10-second ECG segment, averaged over 1000 iterations.

2.4 Summary

In this work, we proposed a novel approach to improve MI diagnosis from ECG signals by incorporating clinical context knowledge of diagnostically salient morphological regions into a deep learning framework. In particular, we developed a cross-task attention transfer framework, which utilized knowledge distillation to distil the ECG morphology knowledge from a teacher network to a student network. The framework integrates a teacher network based on UNet with attention and residual connections, designed specifically for ECG delineation to capture morphological attention on ECG signal. It incorporates a student network

based on ResNet with spatial attention, trained for MI classification, and a CTAT module to facilitate the transfer of morphological knowledge from teacher to student. This knowledge transfer assists the student network in learning more informative and discriminative feature representations for robust MI detection. Experiments on three distinct 12-lead ECG databases demonstrate that our proposed method achieves encouraging performance in MI detection and outperforms other state-of-the-art methods. The proposed CTAT framework provides an advanced methodology for incorporating prior contextual information into deep learning models, specifically for real clinical applications.

3

A Multi-Task Physics-Constrained Hierarchical MI Classification

Contents

In the previous chapter, we discussed an approach for incorporating domain knowledge for MI detection. In this chapter, we build upon this approach to address MI localization. While most approaches in literature have focused on broader classification- determining the presence or absence of MI signatures in ECGs [66, 101], several studies have advanced to incorporate localization of the infarct alongside MI detection [43, 79]. However, such approaches have often considered MI detection and localization as independent tasks and opted for separate models for detection and localization or used a single network with flat classification over MI subtypes and healthy ECGs. While these methods can detect and locate MI concurrently, they overlook the inherent hierarchical relationship between MI detection and its localization. Such a flat classification approach faces inherent challenges due to the complex relationship between detection and location classes, where intra-class variability and inter-class similarity can significantly impact model performance. It treats all categories independently, failing to capture the intrinsic relationship between MI presence and its location while also struggling with class imbalance where certain MI locations may be underrepresented compared to normal cases. This oversight is particularly notable given that in clinical practice, the diagnosis of MI follows a hierarchical process. Additionally, previous deep-learning methods have predominantly relied on a data-driven approach, often utilizing the abstract feature representations extracted at higher layers without considering any domain-specific knowledge as discussed in chapter 2. However, recent research [15, 102] has demonstrated the benefits of integrating relevant clinical heuristics or task-specific context into deep learning frameworks for diagnostic tasks.

To address these issues, we model the MI classification task as a hierarchical classification problem, similar to what is followed in clinical practice, where experts must be able to distinguish between MI and non-MI signals prior to applying methods for localization of the infarcted region. To enable this, we formulate the MI classification task similar to a sequence learning task, where the goal is to find a sequence of labels given a set of input feature representations. Specifically, we employ an encoder-decoder architecture, where a CNN based encoder learns the latent feature representation of the input ECG and an LSTM based decoder generates a sequence of labels iteratively in an autoregressive manner. Here, the sequence refers to the level of classification, where MI detection is the first level and sub-classification of MI forms the second level. For MI localization, we considered five sub-classes of MI: Inferior MI (IMI), Anteroseptal MI (ASMI), Inferolateral MI (ILMI), Anterior MI (AMI), and Anterolateral MI (ALMI) [17]. To incorporate clinical priors corresponding to diagnostically important regions into the model, we leverage a multi-task learning framework. Specifically, we include an auxiliary ECG delineation task, which identifies key regions of the signal known to exhibit pathological changes during MI. This auxiliary task ensures the model attends to these clinically relevant regions during training, thereby enhancing its diagnostic capabilities

for the primary classification task. Additionally, we incorporate an ECG reconstruction task to provide the model with a comprehensive understanding of signal dynamics. This task allows the model to learn the underlying structure and patterns of the ECG signal, further integrating domain-specific knowledge into its representation. To regularize ECG reconstruction from the shared latent space, we additionally incorporated the physics relation among the leads in order to have high-fidelity reconstruction and also to capture the inter-lead information. By jointly optimizing these diverse but clinically relevant auxiliary tasks, we create a synergy that enhances the model's comprehension of ECG morphology and its clinical significance. Essentially, the auxiliary tasks regularize the model to learn robust feature representations, aiding in improved MI classification. The proposed Multi-Task Physics-Constrained Hierarchical MI classification framework (MTPCHi-MI) thereby introduces a novel approach to incorporate clinical context and physics-informed signal dynamics in an end-to-end manner for context-aware hierarchical MI classification.

The remainder of the chapter is organized as follows. Section 3.1 introduces the proposed MTPCHi-MI framework. Section 3.2 describes the experimental setup, while Section 3.3 presents results and analysis. Section 3.4 provides a summary of the chapter.

3.1 Proposed Method: Multi-task physics-constrained hierarchical MI classification

The proposed MTPCHi-MI framework is an end-to-end context-aware hierarchical classification network for improved MI detection and localization from 12-lead ECG signals. Our approach utilizes an MTL framework featuring hierarchical classification of MI as the primary task, with ECG reconstruction and ECG delineation serving as auxiliary tasks. We hypothesize that incorporating these relevant auxiliary tasks can enrich the feature representation, thereby improving the model's context awareness and overall performance. Fig. 3.1 gives the overview of the proposed approach. It follows an encoder-decoder architecture with a shared encoder and task-specific decoders corresponding to MI classification, ECG delineation, and ECG reconstruction. The detailed architecture of each of these networks is presented in section III-B. To model the hierarchical structure within the MI detection and MI localization, we formulate the primary task of hierarchical MI classification as a sequence generation task with an LSTM-based decoder. Additionally, to enforce the hierarchy in the model prediction, we utilized a hierarchical loss function defined by eq. (3.3). All tasks are concurrently optimized, with the ECG reconstruction loss [as presented in (eq. 3.5)] further regularized by the physical laws governing 12-lead ECG signals. This integrated framework, combining MTL, physics-constrained regularization, and hierarchical classification, facilitates comprehensive context-aware feature representation, thereby improving MI detection and localization from ECG signals.

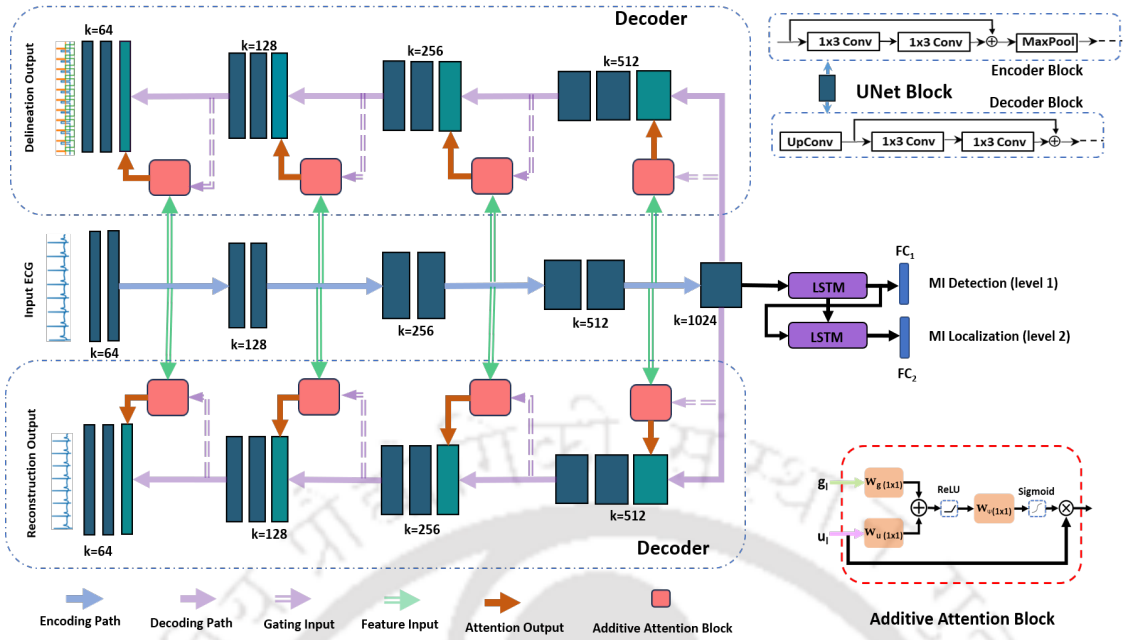

Figure 3.1: Schematic overview of the proposed approach. It features an encoder-decoder architecture with a unified encoder backbone with dual decoding paths for auxilliary tasks and a LSTM based decoder for hierarchical MI classification.

3.1.1 Problem definition

This work aims to improve MI detection and localization from 12-lead ECG signal and seeks to provide a hierarchical granularity to its prediction. Given a dataset $\mathcal{D} = \{(x_n, y_n^c, y_n^d) \mid 1 \leq n \leq N\}$, with N instances where x_n represents an instance, while y_n^c and y_n^d correspond to the ground truths for MI classification and ECG delineation task, respectively. Consider $\mathcal{X} \in \mathbb{R}^{k \times D}$ be input space with k number of leads and D dimensions per lead and $\mathcal{Y}$ denote the label space, where $\mathcal{Y}$ constitutes three parts: $\mathcal{Y}^c$ for hierarchical classification task, $\mathcal{Y}^d$ for ECG delineation task and $\hat{\mathcal{X}}$ for ECG reconstruction task. As such, the objective is to learn a hypothesis per task as $f_t(x; \theta_s, \theta_t) : \mathcal{X} \rightarrow \mathcal{Y}$. where θ_s represents the shared parameters and θ_t refers to task-specific parameters for $t \in T$ and T represents the total number of task ($T = 3$). For an instance x_n in input space $\mathcal{X}$, $y_n^c = (y_{n,1}^c, y_{n,2}^c, ..., y_{n,L}^c)$ represents the vector of L hierarchical labels and for the l^{th} level in the class hierarchy, $y_{n,l}^c \in \{1, 2, ..., V_l\}$ with V_l representing the number of classes within the l^{th} level of the hierarchy. The objective for the primary task of hierarchical classification of MI is to learn to maximize the distribution $p(y_{n,l}^c | x_n, y_{n,1}^c, ..y_{n,l-1}^c)$. In case of ECG delineation task, the objective hypothesis $f^d : \mathcal{X} \rightarrow \mathcal{Y}^d$ with $\mathcal{Y}^d \in \{0, 1\}^{w \times D}$, where w corresponds to the number of ECG wave categories ($w = 3$, corresponding to P wave, QRS complex, T wave). The network is trained by optimizing multiple loss functions corresponding to the primary and the auxiliary tasks, and the contribution of each loss is balanced using a gradient-based approach, which is detailed in sections III-C.

3.1.2 Network Architecture

The network architecture is characterized by a unified encoder backbone that branches into twin parallel decoders. It resembles a bifurcated extension of dual U-Net architecture, having dual decoding paths with a shared encoder. The unified encoder consists of sequential downsampling blocks incorporating residual connection and attention gates. Each block comprises two consecutive 1×3 convolutional layers, each followed by batch normalization and ReLU activation. The encoder progressively downsamples the input through the max pooling layer at each block with a kernel size of 2 while increasing the feature depth. We integrate residual connections into each block to create direct pathways for high-resolution feature propagation to deeper layers. These connections preserve high-resolution spatial details and also ensure efficient gradient flow during training [103]. To ensure compatibility between the dimensions of the residual output and feature maps, a convolutional layer with a kernel size of 1 and a stride of 1 is used. For b^{th} block, the output can be represented as:

$$\mathcal{U}_b = \mathcal{F}(I_{b-1}, \vartheta_b) + \vartheta_b^{proj} * I_{b-1} \tag{3.1}$$

$with,$

$$\mathcal{F}(I_{b-1}) = (ReLU \circ BN \circ conv_{1\times3})^2(I_{b-1}) \tag{3.2}$$

where $\mathcal{F}(\cdot)$ represents the transformation applied by the b^{th} block with parameter ϑ_b to the input to the block, I_{b-1} and $\circ$ denotes the function composition operator. ϑ_b^{proj} denotes the parameters of the convolutional layer that adjusts the input dimensions to ensure compatibility with the identity mapping. The input to b^{th} block, $I_b \in \mathbb{R}^{k_b \times D/2^b}$, $k_b = \{64, 128, 256, 512\}$, $b \in \{1, 2, 3, 4\}$. Here D represents the length of the feature input to the block and k denotes the number of input channels. Each section in the encoder produces multi-scale feature representations that are shared with both decoder branches through skip connections controlled by attention gates.

The shared latent feature representation, $\mathcal{Z} \in \mathbb{R}^{E \times M}$ obtained from the encoder, encapsulates the salient features of the input ECG signal for all of the three tasks. Here, E and M correspond to the channel and temporal dimensions of the latent feature representation. For the classification of MI in a hierarchical manner, we have utilized LSTM-based architecture, drawing inspiration from sequence generation tasks. The latent feature representation, $z \in \mathcal{Z}$, serves as the initial input to a series of L LSTM layers, where L corresponds to the number of levels in the classification hierarchy. Each LSTM layer sequentially processes the output from the previous level, generating a hidden state h^l and cell state c^l. We utilize a temporal attention mechanism to dynamically weight the feature representation, obtaining a

single context vector $\mathcal{C}^l \in \mathbb{R}^E$, which summarizes the weighted information across the temporal dimensions. The temporal attention mechanism processes a sequence of feature vectors $S^l = [s_1^l, ..., s_E^l]$ obtained from the LSTM outputs at each temporal dimension and the hidden state h^l from the final state of the LSTM. For each temporal dimension m, it computes an attention score $e_m = W[h^l; S_m^l] + b$, where $[h^l; S_m^l]$ is the concatenation of h^l and S_m^l, and W and b are learnable parameters of a linear layer. These scores are normalized using softmax: $\beta_m^l = \frac{\exp(e_m)}{\sum_{m=1}^{M} \exp(e_m)}$, producing attention weights that sum to 1. The final context vector is then computed as a weighted sum: $\mathcal{C}^l = \sum_{m=1}^{M} \beta_m^l S_m^l$. This enables the network to selectively focus on the most significant temporal regions based on the attention weights. The context vectors are generated sequentially for l levels in a recursive manner, with each vector being used for a prediction at its corresponding level via linear layers that map the LSTM's output to the corresponding label classes at each level of the hierarchy. The model is trained to minimize the hierarchical cross-entropy loss described in section III-C.

The architecture of twin parallel decoders is symmetrical to the encoder but in reverse order. They utilize bilinear upsampling followed by residual mapping to reconstruct the spatial information. The skip connections from the encoder undergo attention gating before concatenation with the upsampled features in each decoder. The attention mechanism uses convolutional layers and nonlinear activation (ReLU), taking inputs from both the decoder output and features from the corresponding encoding path to learn attention weights, which are subsequently multiplied with the input features, making the model focus on the important regions. For a single decoder branch, let $u^b \in \mathbb{R}^{k_u \times H}$ denote the feature representation from the b^{th} encoder block and $g \in \mathbb{R}^{k_g \times H}$ being the gating input, derived from coarser feature representation from the decoder. Here H are spatial dimensions, and k_u and k_g are channel dimensions of input feature representation and gating input, respectively. The attention coefficients α^i for block i that identifies salient regions in the input feature representation u^i is computed as,

$$\alpha^i = \sigma(q_{att}^i(x^i, g; \Theta_{att})) \tag{3.3}$$

$with,$

$$q_{att}^i = W_\psi^T(ReLU(W_u^T u^i + W_g^T g + b_g))) + b_\psi \tag{3.4}$$

where σ represents to the sigmoid activation and $W_\psi^T \in \mathbb{R}^{k_{int} \times 1}, W_u^T \in \mathbb{R}^{k_u \times k_{int}}, W_g^T \in \mathbb{R}^{k_g \times k_{int}}, b_g \in \mathbb{R}^{k_{int}}, b_\psi \in \mathbb{R}$ are the set of parameters (Θ_{att}) for the attention block, which are optimized during training. The attention coefficients $\alpha^i \in (0, 1)$ are then used to modulate the input feature representation u^l by element-wise multiplication $\hat{u}^i = u^i \cdot \alpha^i$ to obtain the attention block output. Both decoders share the same

attention structure, but each possesses its own independent attention blocks. This configuration allows for shared high-level feature extraction in the encoder while enabling task-specific, attention-driven feature refinement in each decoder path.

3.1.3 Multi-Task Learning Loss

To enforce the network prediction respect the label hierarchies, we adopted the hierarchical cross-entropy [104] based loss function for the primary task. The probability for a given label is calculated by considering the conditional probabilities for each label in the hierarchy. For a class label v in the hierarchy level l, the hierarchical classification loss $\mathcal{L}_{HC}$ is mathematically represented as,

$$\mathcal{L}_{HC}(\theta_s, \theta_{hc}) = -\sum_{l=1}^{L} \lambda(v^l) \log p(v^l|v^{l-1}) \tag{3.5}$$

$with,$

$$p(v^l|v^{l-1}) = \frac{\sum_{i \in v^l} p(i)}{\sum_{j \in v^{l-1}} p(j)} \tag{3.6}$$

where $\lambda(v^l)$ is a weighting factor, which assigns the contribution of each loss term corresponding to the class labels within the label hierarchy and is calculated as $\lambda(v^l) = exp(-\alpha h(v^l)$, with h denoting the level in the hierarchy and $\alpha > 0$, controls the degree to which information is diminished down the hierarchy.

To optimize the task of ECG reconstruction, we adopted Huber loss, which combines the characteristics of Mean Squared Error and Mean Absolute Error. It behaves quadratically for small errors and linearly for large ones, with a transition controlled by hyperparameter β, making it robust to outliers while remaining sensitive to subtle ECG morphology changes. To ensure the network adheres to the underlying physics of ECG, we additionally regularize the reconstruction to comply with Einthoven's Law and the Augmented Leads Sum rule. Mathematically, Einthoven's Law states that $|LeadII - (LeadI + LeadIII)| = 0$. The Augmented Leads Sum rule dictates that $aVR + aVL + aVF = 0$. We incorporate these relations to regularize the reconstruction loss $\mathcal{L}_R$ by adding $\mathcal{L}_{EL}$ (corresponding to Einthoven's Law) and $\mathcal{L}_{ALS}$ (corresponding to the Augmented Leads Sum rule) to the total reconstruction loss function. Given a pair of reconstructed ECG $\hat{x}$ and the target x, we compute the reconstruction loss as:

$$\mathcal{L}_R(\theta_s, \theta_r) = \frac{1}{k} \sum_{k=1}^{K} \mathcal{L}_{Huber}(\hat{x}_k, x_k) + \mathcal{L}_{EL} + \mathcal{L}_{Als} \tag{3.7}$$

and

$$\mathcal{L}_{Huber} = \begin{cases} 0.5(\hat{x}_k - x_k)^2, & \text{if } |\hat{x}_k - x_k| < \beta. \\ \beta(|\hat{x}_k - x_k| - 0.5\beta), & \text{otherwise} \end{cases} \tag{3.8}$$

$$\mathcal{L}_{EL} = ||Lead_{II} - (Lead_I + Lead_{III})||_2^2 \tag{3.9}$$

$$\mathcal{L}_{ALS} = ||aVR + aVL + aVF||_2^2 \tag{3.10}$$

For the ECG delineation task, the network predicts the presence or absence of each constituent wave in the ECG waveform at each time step. To optimize the delineation network, we employ binary cross-entropy loss, which effectively measures the dissimilarity between predicted and true segmentation masks for each constituent ECG wave. Mathematically, it is denoted as,

$$\mathcal{L}_D(\theta_s, \theta_d) = -\sum_{w=1}^{W} (y_w^d \log P_d) + (1 - y_w^d) \log(1 - P_d)) \tag{3.11}$$

where P_d is the prediction probability score and W represents the number of constituent waves. Given the individual task losses, we seek to jointly optimize multiple loss functions during training. Proper balancing of losses is essential to avoid task bias and conflicting gradient directions, which may lead to degraded performance across all tasks. In this work, we adopt PCGrad [105] to properly balance each objective's contribution. In standard multi-task learning, when two task gradients have negative cosine similarity ($\Delta_i \cdot \Delta_j < 0$), they pull the shared parameters in opposing directions, causing one task's learning to partially cancel out the other's progress. PCGrad addresses this by projecting the conflicting gradient onto the normal plane of the other gradient, mathematically: $\Delta_i \leftarrow \Delta_i - \frac{\Delta_i \cdot \Delta_j}{|\Delta_j|^2} \Delta_j$ if $\Delta_i \cdot \Delta_j < 0$, where Δ_i and Δ_j represent the gradients from tasks i and j. This projection removes the component of Δ_i that directly opposes Δ_j while preserving the orthogonal components that can still contribute to task i's learning. As a result, the modified gradients have non-negative cosine similarity, allowing each task to make progress without hindering others. This results in improved overall performance across all objectives without manual tuning of loss weights.

3.2 Experimental Setup

This section describes the experimental methodology used to validate the proposed framework. First, we detail the databases utilized in this study. Next, we present the details of comprehensive comparative

analysis framework comprising ablation studies, quantitative evaluation against existing methods, and qualitative analysis. The evaluation metrics used to assess both the primary and auxiliary tasks are then outlined. Finally, we provide implementation details of the proposed approach, including network architecture specifications, training protocols, and hyperparameter settings.

3.2.1 Datasets

As in Chapter 2, we utilize the *PTB-XL* [91] dataset to train our proposed model. For the MI localization task, we considered the five MI sub-classes (IMI, ASMI, ILMI, AMI, and ALMI). The delineation annotations of ECG waves for each record are acquired from *PTB-XL+* [106] dataset, a supplementary dataset for the PTB-XL. Furthermore, we employ *PTB* [107] diagnostic database, which consists of 549 ECG records from 290 subjects. It has a total of 148 MI subjects with annotation of MI subclasses for each record. We utilized data from PTB database specifically for testing and evaluating our proposed method for both MI detection and localization. In addition, consistent with Chapter 2, we utilize ECG recordings from the *Shandong Provincial Hospital* (SPH) dataset [92] and the *China Physiological Signal Challenge* (CPSC) dataset [108, 109] for the evaluation of MI detection. Both CPSC and SPH datasets serve as the testing dataset for MI detection, facilitating assessment of the proposed method's generalization performance across different datasets.

3.2.1.1 Preprocessing

We follow the preprocessing pipeline described in Chapter 2, where ECG records from all datasets are first preprocessed to remove baseline wander, muscle artifacts, and high-frequency noise. A moving average filter [110] is applied to eliminate baseline wander, followed by a non-local means (NLM) filter [94] to suppress muscle artifacts and high-frequency noise. The resulting signals are then normalized using Z-score normalization and further assessed for quality based on the criteria outlined in [95]. For a consistent sampling rate across databases, all the signals are resampled to 250 Hz. Furthermore, 10-sec segments are extracted from all the records to maintain uniform segment length across the databases. For the task of ECG delineation, binary sequences corresponding to P-waves, QRS complexes, and T-wave annotations were derived for each segment. These sequences served as the ground truth for the ECG delineation task. During training, we employ data from train folds (1–8) of the PTB-XL database. Fold 9 of the PTB-XL database acts as the validation set. We utilize the test fold of PTB-XL to evaluate both MI detection and localization performance. Meanwhile, we reserve the SPH and CPSC datasets exclusively for testing MI detection performance.

3.2.2 Comparative Analysis

To empirically validate the efficacy of the proposed method, we conducted multiple groups of comparative experiments. The details of the experiments are as follows.

- **Ablation Study**: We first demonstrate the effectiveness of our proposed MTL framework compared to a single-task counterpart for the primary task. We trained a model without auxiliary tasks for hierarchical MI classification as a signal task, which acted as a benchmark model for comparison (*baseline*). Furthermore, an investigation is conducted to compare the PCgrad approach of multi-task optimization against the equal weight loss strategy in the context of MTL. Additionally, we assess the influence of each auxiliary task by integrating each into the baseline. First, we incorporate the decoder network of the ECG delineation Task (EDN), and subsequently, we investigate the influence of physics-constrained ECG reconstruction task (PCERN) in the performance of the classification network. To validate the effectiveness of physics-constrained ECG reconstruction, we compared our approach with an identical ECG reconstruction network (ERN) without physics-constrained regularization.
- **Qualitative Analysis**: For a better understanding of the predictions made by the network, we plot saliency map [111], which highlight the most influential regions of the input ECG contributing to the final prediction. Specifically, these maps are generated by computing the gradient of the predicted class score with respect to each time point in the input ECG signal. This gradient-based visualization allows us to identify the temporal segments of the ECG that the model considers most critical for classification.
- **Quantitative Comparison**: We compare our proposed approach with five deep learning-based state-of-the-art methods of MI classification from multi-lead ECG signals: Spa-Tem MI [43], ML–ResNet [67], CTAT [15], MFB-CBRNN [66], and MCDANN [112], also with two feature-engineering-based methods: IMET-T2D [113] and MEES [38]. While CTAT and MFB-CBRNN are methods focused on MI detection, we extended these approaches to incorporate multiclass classification to perform MI localization. To ensure fair comparisons, we implemented each method following our experimental setup. Each method was trained using PTB-XL train folds and evaluated under identical experimental conditions.

3.2.3 Evaluation Methods

To evaluate the fidelity of the reconstructed ECG, we selected four different metrics: Percentage root-mean-square difference (PRD), Dynamic time warping (DTW), Wavelet Energy based diagnostic distortion

(WEDD) [114], and Structural similarity index (SSIM) [115], each designed to assess a specific aspect of the reconstruction. PRD is a standard distortion metric commonly used to evaluate ECG reconstruction by quantifying the percentage root mean square difference between the original and reconstructed signals. WEDD is a noise-robust diagnostic distortion measure that compares the wavelet energy distributions. DTW evaluates the temporal alignment of the signals, and SSIM assesses the structural similarity by comparing local windows for structural information, such as the shape and timing of different constituent waves. The delineation performance is evaluated using onset and offset F1 score, precision, recall and relative timing error at tolerance (TOL) of 25, 70, and 150 ms for each constituent wave. Tolerance refers to the time window within which the onset/offset prediction is deemed correct. The relative timing error is computed by taking the mean and standard deviation ($\mu \pm \sigma$) of the difference between the actual and predicted onsets/offsets, considering instances where both the predicted and ground truth labels are within a specific tolerance window.

To evaluate the model performance for MI classification, we employ two sets of metrics: one to evaluate level-wise performance on MI detection and localization and the second to evaluate the overall hierarchical structure of the classification. We utilize the F1 score, AUROC [97], and PRAUC [98] to evaluate the level-wise MI detection and localization. To measure the hierarchical performance, we employ Hierarchical F-Measure (F_H) [116], Tree Induced Error (TIE) [117] and overall accuracy (OA). OA is the accuracy measure considering each unique combination of hierarchical labels as a distinct class. It represents the proportion of instances where the predicted label combination exactly matches the true label combination. TIE measures the hierarchical distance between predicted and true classes, assigning costs to misclassifications based on their level in the hierarchy, with higher-level errors incurring greater penalties. A high TIE suggests the model struggles with the hierarchical structure, leading to misclassifications from error propagation.

3.2.4 Implementation Details

The implementation of the proposed method was carried out in Python using PyTorch (v2.0.1), and the experiments were conducted on an Ubuntu-based workstation equipped with an Intel(R) Xeon(R) Silver 4214R 32-core CPU @ 2.40 GHz, 125 GiB of RAM, and an NVIDIA A100-PCIE GPU. The development environment was Visual Studio Code. Training was performed using the SGD optimizer with a momentum of 0.99 and an initial learning rate of 0.001, using a mini-batch size of 32. To reduce the risk of overfitting, a learning rate scheduler was employed, which decreased the learning rate by a factor of 5 when no improvement in validation loss was observed over 10 consecutive epochs. Furthermore, early stopping

was applied to terminate training if validation loss failed to improve over 25 consecutive epochs. The model that achieved the lowest validation loss during training was selected for subsequent analysis.

3.3 Results and Discussions

In this section, we evaluate the performance of our proposed framework for hierarchical MI classification along with the auxiliary tasks of ECG reconstruction and ECG delineation. For training, the ECG signals from the PTB-XL dataset are preprocessed using the preprocessing pipeline described in section IV-B. The signals thus obtained are split into the train (fold 1-8), validation (fold 9), and test (fold 10) sets, as per database split recommendations in [91]. The training set consisted of 8179 ECG signals, while the validation set comprised 1455 signals. For ECG delineation, the ground truth annotations for the P-waves, QRS, and T-waves were obtained from *PTB-XL+* [106] dataset.

3.3.1 Performance Analysis of Auxiliary tasks

3.3.1.1 Results on ECG reconstruction

In this section, we assess the ECG reconstruction performance of our proposed network. We evaluate various aspects of reconstruction quality using four distinct metrics described in the section 3.2.1.1. Table 3.1 summarizes the reconstruction results across all 12 ECG leads. The overall results demonstrate high-fidelity reconstruction across all the leads. Specifically, $lead - V5$ demonstrated the lowest WEDD of 0.1198, indicating diagnostic features are well preserved for this lead. It can be observed that the limb leads (I, II, III, aVR, aVL, aVF) exhibited higher PRD values compared to the precordial leads (V1-V6). For example, Lead III showed the highest PRD of 0.2660, indicating more significant reconstruction challenges for this particular lead. In contrast, Lead V5 demonstrated the lowest PRD of 0.1305, suggesting superior reconstruction fidelity among the precordial leads. Regarding SSIM, we observed that Lead V4 had the highest value at 0.9236, while Lead III showed the lowest at 0.8514. This pattern aligns with our PRD findings, reinforcing that our network generally achieves better reconstruction for precordial leads. These reconstruction quality differences can be attributed to the inherent signal characteristics of limb versus precordial leads. Precordial leads capture localized cardiac electrical activity from specific myocardial regions with consistent waveform morphology across subjects, while limb leads measure potential differences between extremity electrodes and represent integrated cardiac electrical activity with greater inter-subject variability, higher susceptibility to motion artifacts and noise, and typically lower signal amplitudes. Additionally, Figure 3.2 shows examples of ECG reconstructions for various leads compared to the original signal. These visual comparisons provide a qualitative complement to our quantitative metrics,

allowing for a more comprehensive assessment of our network's reconstruction capabilities across different lead types.

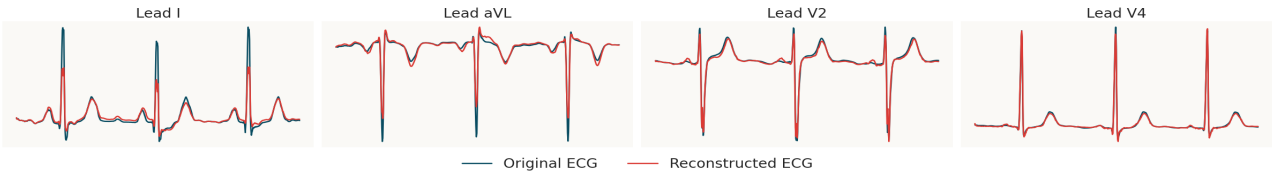

Figure 3.2: Comparison between the reconstructed and original signals for Lead I, Lead aVL, lead V2 and Lead V4.

Table 3.1: ECG Reconstruction Performance across PTB-XL Test Set and CPSC Dataset. Vertical arrows denote the direction of desirable performance.

		Leads												
Database	Metrics	I	II	III	aVL	aVR	aVF	V1	V2	V3	V4	V5	V6	Mean ± Std
PTB-XL (Test-Fold)	PRD ↓	0.2125	0.2286	0.2660	0.2071	0.2556	0.2390	0.1412	0.1394	0.1596	0.1351	0.1305	0.1350	0.1875 ± 0.0501
	DTW ↓	7.9051	8.4643	10.1191	7.6588	9.7103	9.1059	5.0570	4.9761	5.6310	5.1769	4.8879	4.8114	6.9587 ± 1.9833
	WEDD ↓	0.2030	0.2200	0.2577	0.1972	0.2445	0.2300	0.1297	0.1281	0.1478	0.1214	0.1198	0.1256	0.1771 ± 0.0510
	SSIM ↑	0.9033	0.9084	0.8514	0.9168	0.8640	0.8920	0.8918	0.8926	0.9017	0.9236	0.9215	0.9088	0.8980 ± 0.0209
CPSC	PRD ↓	0.2432	0.2325	0.2464	0.2255	0.2706	0.2289	0.1759	0.1650	0.1896	0.1634	0.1810	0.1886	0.2092 ± 0.0344
	DTW ↓	9.1956	8.4152	9.6229	7.9494	10.6309	8.5241	5.7924	5.4412	6.2604	5.4780	5.6737	5.8781	7.4052 ± 1.7774
	WEDD ↓	0.2353	0.2235	0.2364	0.2141	0.2614	0.2175	0.1693	0.1557	0.1799	0.1533	0.1697	0.1803	0.1997 ± 0.0343
	SSIM ↑	0.8300	0.8628	0.8132	0.8558	0.8010	0.8486	0.8342	0.8439	0.8488	0.8733	0.8505	0.8387	0.8417 ± 0.0193

3.3.1.2 Results on ECG Delineation

We evaluate the ECG delineation performance of the proposed network by assessing its ability to accurately identify the onset and offset of P waves, QRS complexes, and T waves. Table 3.2 summarizes the waveform delineation performance across the PTB XL + test set. We evaluated predictions using three TOL windows: 25, 70, and 150 ms. These windows define the acceptable margin of error for predictions. A prediction is deemed accurate if it falls within the specified TOL window of the actual boundary. Results show that the QRS complex boundaries were most accurately detected, with an F1 score of 98.33 % for onset and 98.45 % for offset at a TOL of 150ms. P wave delineation showed the next best performance, with F1 scores of 95.12 % and 95.49 % for onset and offset, respectively. T wave delineation proved to be the most challenging, with F1 scores of 88.77 % for onset and 89.51 % for offset. This relatively lower performance is likely due to the T wave's variable morphology and changes in the ST segment, as seen in

cases of myocardial ischemia, which can obscure the boundary between the end of the ST segment and the beginning of the T wave. The performance pattern remains consistent across all tolerance windows. Figure 3.3 presents a visual illustration of the delineation results, highlighting the detected boundaries across the 12-lead ECG signals with diverse morphological patterns. These figures illustrate the network's capability performance across diverse ECG morphologies, providing visual evidence that complements our quantitative metrics.

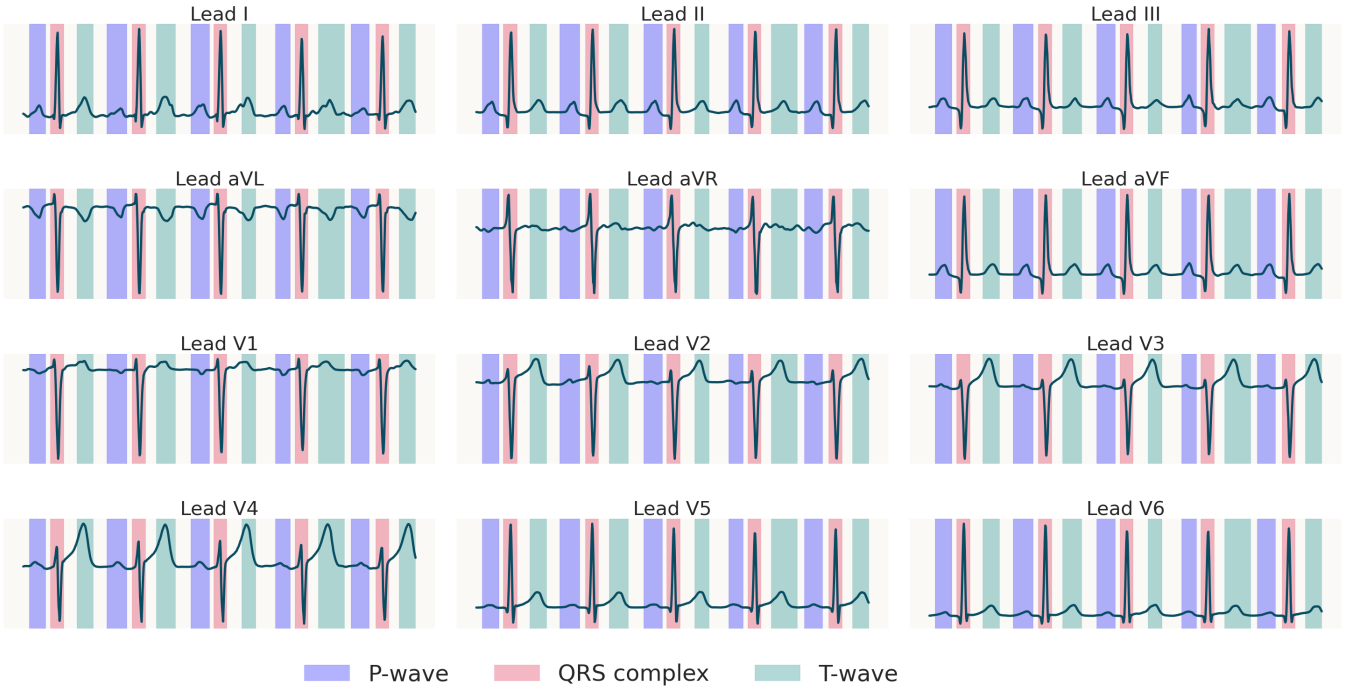

Figure 3.3: Model predictions for P-wave, QRS complex, and T-wave delineation across the 12-Leads.

Table 3.2: ECG delineation performance on PTB-XL+ Test Set

ECG waves	TOL(ms)→	Precision ↑			F1-Score ↑			Error($\mu \pm \sigma$) ↓			Mean IOU ↑
		25	70	150	25	70	150	25	70	150	
P	Onset	88.98	90.40	90.72	94.14	94.94	95.12	1.21 ± 1.40	2.68 ± 3.98	3.46 ± 5.66	0.9139
	Offset	90.68	91.28	91.41	95.09	95.42	95.49	0.84 ± 1.17	1.59 ± 3.06	1.95 ± 4.12	
QRS	Onset	96.68	96.76	96.76	98.28	98.32	98.33	0.60 ± 1.03	0.81 ± 1.67	0.81 ± 1.70	0.9629
	Offset	96.87	97.00	97.01	98.37	98.44	98.45	0.75 ± 1.24	1.15 ± 2.27	1.17 ± 2.36	
T	Onset	77.07	79.08	79.82	87.04	88.31	88.77	1.03 ± 1.22	2.29 ± 3.88	3.18 ± 5.69	0.8913
	Offset	76.91	79.49	81.02	86.93	88.56	89.51	0.86 ± 1.13	2.58 ± 4.53	4.53 ± 7.68	

3.3.2 Results on Hierarchical MI Classification

Table 3.3 provides the quantitative performance of the proposed method for hierarchical MI classification. For MI detection, the model effectively distinguishes between healthy and MI cases, with high F1 scores, AUPRC, and AUROC values of 92.45 %, 97.55 % and 96.05 % respectively. To evaluate the MI localization performance, we consider the records that have non-healthy ground truth and prediction at the second step of classification. It is a multi-class classification over five subcategories of MI (IMI, ASMI, ILMI, AMI, and ALMI). We observe a gradual decrease in performance metrics with an F1 score of 80.81% and PRAUC of 88.16 %. This trend can be attributed to varying degrees of difficulty in distinguishing between these subcategories, particularly among the classes with similar infarction regions, such as ASMI and AMI, or ILMI and IMI, as evident in the confusion matrix (presented in Fig. 3.5). ASMI and AMI both affect the anterior wall of the left ventricle, with ASMI specifically involving the septal region, while ILMI and IMI involve the inferior wall, with ILMI extending to the lateral wall. The hierarchical performance metrics provide insight into the model's effectiveness across the overall classification structure. The hierarchical F1 score of 92.67 % demonstrates strong overall performance, balancing precision and recall effectively across both levels of the hierarchy, while the overall accuracy of 85.59 % shows that a significant majority of predictions are correctly classified considering both levels of the hierarchy. However, the slight gap between accuracy and the F1 score, along with a TIE of 0.4397, suggests that while the model occasionally makes errors, these mistakes typically adhere to the hierarchical structure, such as correctly predicting the broader category but misclassifying the specific subclass. Similar performance was observed on PTB dataset for MI detection with with F1 score of 87.74%, PRAUC of 98.56%, and AUROC of 94.38%. As observed in Table 3.3, the model's performance decreases for MI localization tasks, primarily due to misclassifications between related MI locations, particularly between anterior subtypes (AMI, ASMI, ALMI) and inferior subtypes (IMI, ILMI), consequently affecting the hierarchical classification performance. To validate this observation, we recalculated the model's performance for PTB database using combined anatomical regions as labels and obtained an F1 score of of 88.25%, PRAUC of 85.04%, and AUROC of 91.62%. These results highlight that despite the subtle challenges in fine-grained classification, the model effectively identifies the primary anatomical locations of MI, maintaining strong diagnostic performance at the regional level for this independent dataset.

In Fig 3.4, we visualize the attribution map highlighting the relative importance of specific segments across the input ECG for the model's prediction. We employed Saliency maps [111] to produce attribution maps having a similar structure as of input ECG. We visualize the normalized saliency scores superimposed

as heat maps for a 3-s segment of ECG signal for all the 12 leads of a subject with MI diagnosis within the PTB-XL test set. It can be observed that the model consistently emphasized certain regions across multiple leads, particularly around the QRS complex and ST segment, which are critical in the clinical assessment of MI. For example, in leads III and aVF, increased saliency was observed around the ST elevation, a key indicator of MI severity. Similarly, in lead I and II, the model focused on the T wave's morphology, which is often associated with ischemic changes. The saliency map visualizations validate the model's focus on clinically significant ECG segments and provide a tool for understanding model prediction.

Table 3.3: Hierarchical MI classification performance of the proposed network on PTB XL Test Fold

Database			Predicted (%)		MI Detection			MI localization			Hierarchical		
			HC	MI	F1 Score ↑	AUROC ↑	PRAUC ↑	F1 Score ↑	AUROC ↑	PRAUC ↑	TIE ↓	F_H ↑	OA ↑
PTB-XL	Actual	HC	93.15	6.85	92.45	97.56	96.05	80.81	93.92	88.16	0.4397	92.67	85.59
		MI	8.93	91.07									
PTB	Actual	HC	96.76	3.23	87.74	94.38	98.56	50.66	78.03	50.73	1.1884	80.19	53.70
		MI	15.85	84.15									

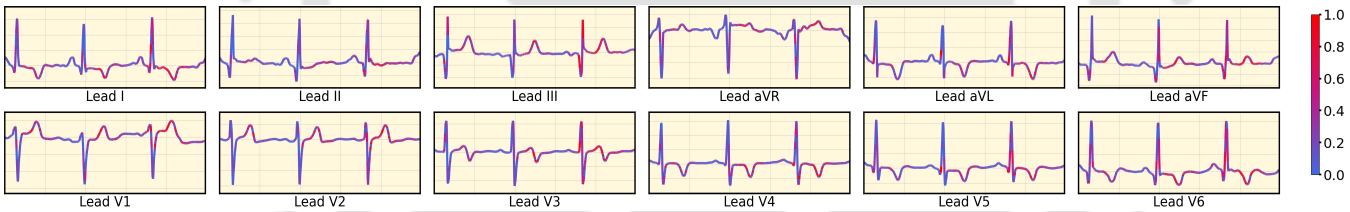

Figure 3.4: Visualization of saliency maps obtained by the proposed method, overlayed over corresponding ECG signal for a subject with MI diagnosis from PTB-XL Test Set.

Table 3.4: Comparative results between single and multi-task learning approaches on PTB XL Test Set

Method	MI Detection			MI localization			Hierarchical		
	F1 Score ↑	AUROC ↑	PRAUC ↑	F1 Score ↑	AUROC ↑	PRAUC ↑	TIE ↓	F_H ↑	OA ↑
Single Task	90.83	97.11	95.35	71.12	90.91	81.63	0.5372	91.05	82.41
Equal weights MTL	91.88	97.71	96.35	76.76	92.78	86.21	0.4655	92.24	84.91
This Work	**92.45**	**97.56**	**96.05**	**80.81**	**93.92**	**88.16**	**0.4397**	**92.67**	**85.59**

3.3.3 Ablation Study

Table 3.4 compares single-task and multi-task learning approaches. The results confirm the superiority of our method compared to single-task network and equal weight-based MTL approach. It can be observed

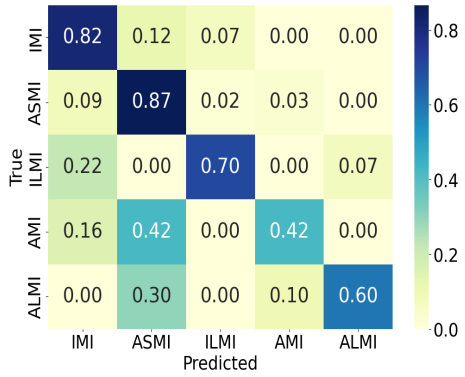

Figure 3.5: Confusion matrix for MI localization

that the single task network achieves an F1 Score of 90.83 % and 71.12 % for MI detection and MI localization, respectively. For the hierarchical classification, it attains a TIE of 0.5372, F_H of 91.05 % and OA of 82.41%. The proposed multi-task network improves the F1 Score by 1.62% for MI detection and by 9.69 % for MI localization. Additionally, the proposed network also improves the performance across all the hierarchical metrics compared to the single task model. These improvements highlight the feature-sharing aspect of multi-task learning. A comparison with the loss addition-based multi-task approach is also performed to evaluate the effectiveness of multi-objective optimization of PCgrad. As shown in Table 3.4, our approach employing PCgrad outperforms the equal-weights loss addition counterpart across all metrics for the overall hierarchical MI classification.

We conduct an ablation study to assess the impact of the auxiliary tasks on the performance of the primary task. Table 3.5 shows the comparison results of incorporating each auxiliary task to the baseline. In comparison to the baseline, the addition of EDN improves the F1 score by 1.24% for MI detection and by 5.01% for MI localization. This indicates that the addition of ECG delineation as an auxiliary task is beneficial for MI classification, as it makes the network aware of the morphological structure of the ECG. Similar performance improvements were observed with the incorporation of PCERN for signal reconstruction as an auxiliary task, suggesting that learning to reconstruct the signal adds additional context about the signal dynamics, which leads to improved MI classification performance. It can also be noted from Table 3.5 that the addition of both auxiliary tasks leads to superior performance compared to incorporating either one alone and results in improvement of all metrics compared to baseline. A comparison with ECG reconstruction without physics-constrained regularization was also performed. In this approach, the reconstruction network (ERN) was trained without the physics-constrained losses. As shown in Table 3.5, the network with PCERN outperforms ERN for overall MI classification tasks. For ECG reconstruction specifically, ERN achieved a mean SSIM value of 0.8899 on the PTB-XL test set, relatively

lower than the proposed method's values of 0.8980, demonstrating the effectiveness of physics-constrained regularization. These findings confirm that the inclusion of auxiliary tasks contributes to more effective feature representations aware of morphological context and signal dynamics of ECG, leading to enhanced overall performance for MI classification.

Table 3.5: Ablative study of different components in our proposed method

	MI Detection			MI localization			Hierarchical		
Method	F1 Score ↑	AUROC ↑	PRAUC ↑	F1 Score ↑	AUROC ↑	PRAUC ↑	TIE ↓	F_H ↑	OA ↑
Baseline	90.83	97.11	95.35	71.12	90.91	81.63	0.5372	91.05	82.41
Baseline + EDN	92.07	97.46	95.97	76.03	93.33	87.61	0.4654	92.24	84.64
Baseline + ERN	91.88	97.13	95.61	73.49	90.45	82.14	0.4709	92.15	84.56
Baseline + PCERN	92.01	97.50	96.10	77.82	93.43	86.37	0.4696	92.17	84.57
This Work	**92.45**	**97.56**	**96.05**	**80.81**	**93.92**	**88.16**	**0.4397**	**92.67**	**85.59**

3.3.4 Quantitative Comparison

We additionally validate our approach by comparing its performance against five state-of-the-art deep learning methods for ECG-based MI detection and localization: Spa-Tem MI [43], ML–ResNet [67], CTAT [15], MFB-CBRNN [66], and MCDANN [112]. The results are reported on the PTB XL test for MI detection and localization, and additionally, we employ the SPH dataset and CPSC dataset specifically to evaluate MI detection performance. As evidenced in Table 3.6, our method demonstrates superior performance in MI detection for the PTB XL and SPH datasets while achieving competitive results for the CPSC dataset. Additionally, our approach demonstrates the highest efficacy for MI localization compared to other methods. CTAT [15] leverages attention distillation to incorporate domain knowledge of MI-induced ECG morphological variations through morphological attention. This approach achieves F1 scores of 92.23 % for MI detection and 78.81 % for MI localization on the PTB XL dataset. Spa-Tem [43] employed self-attention to capture spatio-temporal ECG patterns, whereas MCDANN [112] utilized squeeze-and-excitation based channel attention to obtain heartbeat features. These attention-based approaches achieved competitive performance for both MI detection and localization across the datasets. However, our approach outperforms these methods owing to additional morphological context, and physics-constrained signal dynamics learned through joint optimization of ECG delineation and reconstruction tasks. Furthermore, the modelling of the class hierarchy in the MI classification task has added additional semantic correlation of MI and its subcategories, contributing to the overall superior performance of our method. The performance of conventional feature engineering-based methods was relatively lower compared to

deep learning methods. The performance gap can be attributed to their dependence on manually designed features, while deep learning models automatically learn complex patterns directly from raw ECG signals that are difficult to encode through traditional feature engineering.

Table 3.6: Comparison Results of our proposed method and other existing state-of-the-art methods for MI Classification

	MI Detection									MI Localization		
	PTB-XL (Test fold)			SPH			CPSC			PTB-XL (Test fold)		
Method	F1 Score ↑	AUROC ↑	PRAUC ↑	F1 Score ↑	AUROC ↑	PRAUC ↑	F1 Score ↑	AUROC ↑	PRAUC ↑	F1 Score ↑	AUROC ↑	PRAUC ↑
Spa-Tem MI [43]	92.10	97.04	95.75	96.04	99.13	84.08	91.40	97.38	98.43	78.42	91.74	83.97
ML–ResNet [67]	90.38	96.34	94.58	95.07	98.54	82.69	91.61	97.56	98.58	76.23	92.88	86.46
MFB-CBRNN [66]	90.82	97.09	95.45	95.88	98.90	82.49	91.77	97.53	98.51	76.59	91.87	85.05
MCDANN [112]	90.04	96.79	95.17	95.47	98.89	82.87	91.43	97.51	98.59	74.21	93.46	87.16
CTAT [15]	92.23	97.29	95.79	**96.69**	99.39	87.13	91.97	**97.76**	98.63	78.81	92.83	87.16
MEES [38]	80.10	87.93	82.01	90.62	90.13	61.66	77.19	86.67	90.99	62.33	81.54	70.08
IMET-T2D [113]	84.63	91.28	87.43	91.51	93.21	71.21	83.86	91.19	94.22	69.75	84.76	73.61
This Work	**92.45**	**97.55**	**96.05**	96.39	**99.40**	**87.90**	**92.26**	97.74	**98.65**	**80.81**	**93.92**	**88.16**

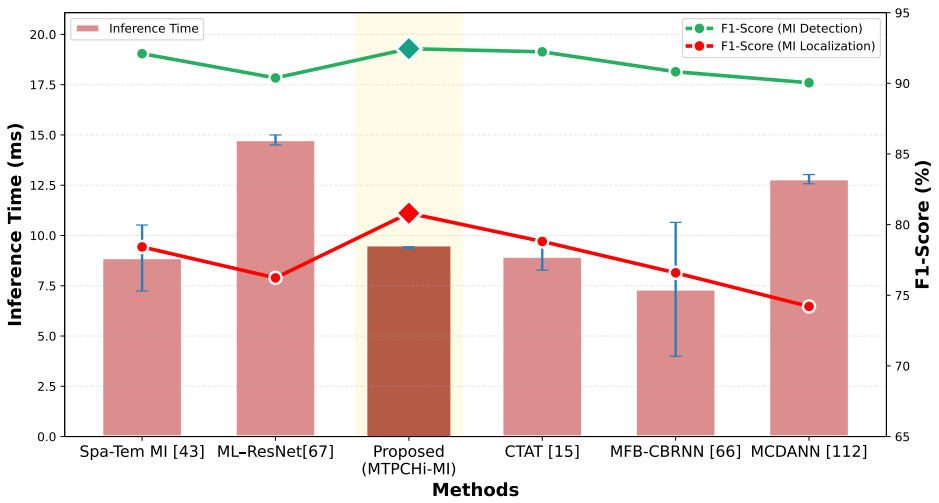

Figure 3.6: Comparison of inference time and classification performance across methods on PTB-XL test set. Bars represent mean inference time per 10-second ECG record with standard deviation (left y-axis). Green and red lines indicate F1 scores for MI detection and localization, respectively (right y-axis).

To evaluate the computational efficiency of the proposed method, we conducted a comparative analysis of inference time across all state-of-the-art-methods, as illustrated in Fig. 3.6. All measurements were performed on an NVIDIA A100 GPU over 1000 iterations to ensure fair comparison. During inference, our implementation leverages only the shared encoder and classification decoder pathways, as the auxiliary delineation and reconstruction branches serve exclusively as training mechanisms to guide the learning of enriched feature representations and are not required for prediction. The proposed method achieves

an average inference time of 9.38±0.06 ms per 10-second ECG record, which is highly competitive with other methods. Specifically, while methods such as Spa-Tem MI [43] (8.5 ± 0.2 ms) and CTAT [15] (8.95 ± 0.68 ms) demonstrate slightly faster inference, and MFB-CBRNN [66] (7.2 ± 3.8 ms) shows the lowest mean latency, our method maintains comparable efficiency. Notably, ML-ResNet [67] (14.5±0.5 ms) and MCDANN [112] (12.8 ± 0.5 ms) exhibit substantially higher inference times compared to our approach. As depicted in Fig. 3.6, the proposed method achieves superior classification performance with F1 scores of 92.45% for MI detection and 80.81% for MI localization, representing the highest performance among all methods. This demonstrates that our multi-task learning framework achieves an optimal balance between diagnostic accuracy and computational efficiency. The inference speed remains well within acceptable limits for real-time clinical decision support systems, making the approach practical for deployment in clinical settings where both accuracy and efficiency are critical.

3.4 Summary

In this chapter, we proposed a unified multi-task framework, MTPCHi-MI, for context-aware hierarchical classification of MI. Specifically, we employ auxiliary tasks of ECG delineation and ECG reconstruction to enable a more complex shared latent representation of ECG data, capturing additional morphological context along with physics-informed signal dynamics. In particular, we employ an encoder-decoder architecture with a shared encoder that learns the latent feature representation common across the three tasks. The architecture features twin parallel decoders based on U-Net with attention and residual connections for the auxiliary tasks and an LSTM-based decoder for hierarchical MI classification. The autoregressive nature of the LSTM decoder allows the model to make predictions at each level of the classification hierarchy in an iterative manner. The hierarchical loss further enforces the network prediction to respect the label hierarchies by leveraging the correlation within MI detection and localization. Experiments on four distinct 12-lead ECG databases demonstrate that our proposed method achieves encouraging performance in MI detection and localization in a hierarchical manner and outperforms other state-of-the-art methods. The proposed MTPCHi-MI network presents a novel deep-learning approach for ECG analysis, enhancing MI detection and localization from 12-lead signals and potentially advancing clinical ECG interpretation.

4

Uncertainty-Aware Multi-Label ECG Classification

Contents

The rapid advancement of deep learning in ECG analysis has demonstrated remarkable potential in automating cardiac diagnosis, offering high accuracy and efficiency in detecting various cardiac conditions. However, the deployment of these systems in real-world clinical settings faces crucial challenges that could compromise patient safety and diagnostic reliability [118]. Deep learning models, while powerful in classification tasks, often provide point estimates without quantifying their predictive uncertainty, potentially leading to overconfident predictions, even when the predictions are wrong [119]. These models generally struggle with out-of-distribution (OOD) data, where test samples deviate significantly from the training data distribution or exhibit disease characteristics not represented in the training set [120]. This limitation becomes particularly critical in healthcare applications, as encountering patients with anomalous or atypical patterns is common in clinical practice, and misdiagnosis can have severe consequences. For instance, a model trained to detect specific cardiac conditions may generate overconfident but incorrect predictions when encountering ECG patterns representing cardiac abnormalities that were not included in its training data. Furthermore, traditional deep learning models typically provide only point predictions without quantifying their predictive confidence, treating both straightforward and challenging cases equally. This limitation becomes particularly problematic in clinical practice, where understanding prediction confidence is crucial for determining when additional diagnostic evaluations may be necessary. Without reliable uncertainty estimates, healthcare providers lack the necessary information to assess the trustworthiness of automated diagnoses, potentially leading to suboptimal clinical decisions.

In the previous chapters, we presented the proposed automated approaches for detection of MI and its subtypes from 12-lead ECG signals. However, in clinical practice patients may manifest multiple cardiac conditions, potentially occurring simultaneously and with diverse pathophysiological origins. For Such cases, the model is required to perform multi-label classification. In this approach, the system must independently assess the presence or absence of each potential cardiac condition from a predefined list of diagnoses. This differs from multi-class classification, where a single class label is assigned per input. In this study, we extend the classification task to incorporate five diagnostic labels in a multi-label framework, namely Myocardial Infarction (MI), ST/T Changes (STTC), Conduction Disturbance (CD), Hypertrophy (HYP), and Normal ECG (NORM). Previous approaches for multi-label classification have either followed binary relevance [70], where separate binary classifiers are trained for each label independently, or label powerset [121, 122] methods that transform the multi-label problem into a multi-class problem by considering each unique combination of labels as a separate class. However, these methods often fail to capture the complex interdependencies between different cardiac conditions, as they treat class predictions independently of each other. In clinical practice, there exist strong correlations among various

cardiac abnormalities, and leveraging this correlation information among the labels can aid in more precise diagnosis.

This chapter introduces a novel uncertainty-aware deep learning framework for multi-label ECG diagnosis. The framework addresses two key clinical challenges: a representation learning method that captures the inherent correlations between different cardiac conditions, and an uncertainty quantification approach for robust OOD detection to identify ECG patterns where reliable predictions cannot be made. The remainder of this chapter is structured as follows: the proposed approach is detailed in Section 4.1, experimental procedures are outlined in Section 4.2, results and analysis are presented in Section 4.3, and Section 4.4 offers the summary of the chapter.

4.1 Proposed Method: Uncertainty-Aware Multi-Label Cardiac Diagnosis Framework

Taking inspiration from clinical practice, where cardiologists seek second opinions when facing diagnostically challenging cases, automated ECG interpretation systems should similarly recognize and communicate their diagnostic uncertainty. To address this critical need, we propose an uncertainty-aware framework for diagnosis of multi-label cardiac abnormalities. Our framework builds upon variational inference principles, specifically employing a Gaussian Mixture Model Variational Autoencoder (GMM-VAE) architecture that enables probabilistic modeling of the complex ECG signal distribution. Variational inference provides a principled approach to approximate complex posterior distributions, making it particularly suitable for capturing uncertainties inherent in ECG interpretation. Specifically, we design the latent space with K gaussian components matching the number of cardiac conditions in our label space, establishing a one-to-one correspondence between mixture components and diagnostic labels. To fully leverage this structured representation, we introduce a multivariate sigmoid activation that replaces the conventional sigmoid function used for component probability estimation. This correlation-aware mechanism incorporates a learnable interaction matrix that explicitly models the physiological relationships between cardiac conditions, allowing the model to capture both co-occurrence patterns and mutual exclusivity constraints observed in clinical practice. By enforcing label coherence, this structured probabilistic framework improves multilabel classification performance while providing precise uncertainty quantification that differentiates between novel ECG patterns and boundary ambiguity cases. The resulting probability distribution naturally accommodates the multi-label nature of ECG interpretation while providing a robust foundation for uncertainty quantification and out-of-distribution detection in clinical diagnostic settings.

4.1.1 Preliminaries

The Variational Autoencoder [123] is a generative model that learns to map data x to a lower-dimensional latent space z through an encoder-decoder architecture while maintaining a probabilistic interpretation. Unlike traditional autoencoders which map inputs to specific points in the latent space, variational autoencoders learn to model probability distributions over possible latent representations [124, 125]. The encoder, parameterized by ϕ, approximates the posterior distribution $q_\phi(z|x)$, while the decoder, parameterized by θ, models the likelihood $p_\theta(x|z)$.The VAE is trained by maximizing the Evidence Lower Bound (ELBO): $\mathcal{L}_{VAE} = \mathbb{E}q_\phi(z|x)[\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x)||p(z))$ where the first term represents the reconstruction quality and the second term is the Kullback-Leibler divergence between the approximate posterior and the prior $p(z)$. The approximate posterior is typically parameterized as a Gaussian distribution: $q_\phi(z|x) = \mathcal{N}(\mu_\phi(x), \sigma_\phi(x))$. The GMM-VAE extends the standard VAE by replacing the simple Gaussian prior with a mixture of K Gaussian components. The prior becomes: $p(z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(z|\mu_k, \Sigma_k)$ where π_k are the mixture weights, and μ_k, Σ_k are the parameters of each Gaussian component. For our cardiac diagnosis framework, K is set to match the number of possible cardiac conditions, establishing a direct correspondence between mixture components and diagnostic labels. The modified ELBO for GMM-VAE becomes: $\mathcal{L}_{GMM-VAE} = \mathbb{E}q_\phi(z|x)[\log p_\theta(x|z)] - D_{KL}(q_\phi(z|x)||\sum_{k=1}^{K} \pi_k \mathcal{N}(z|\mu_k, \Sigma_k))$.

4.1.2 Problem Definition

Given an input ECG signal $x \in \mathbb{R}^d$ and a label space $\mathcal{Y} = \{0, 1\}^5$ representing the presence/absence of five cardiac conditions, our objective is to learn a mapping function $f : \mathbb{R}^d \to \{0, 1\}^5$ that predicts multiple cardiac conditions simultaneously while providing uncertainty estimates and reliable confidence scores. Formally, for each input x, we aim to:

- Predict a binary vector $\hat{y} \in \{0, 1\}^5$ where $\hat{y}_i = 1$ indicates the presence of the i-th cardiac condition.
- Detect out-of-distribution samples through a decision function $g(x)$: $g(x) = \begin{cases} 1, & \text{if } x \text{ is in-distribution} \\ 0, & \text{if } x \text{ is OOD} \end{cases}$

The model employs a GMM-VAE with five Gaussian components in the latent space, corresponding to the five possible cardiac conditions. The encoder network maps input ECG signals x to latent distribution parameters $q_\phi(z|x) = \prod_{k=1}^{5} \mathcal{N}(\mu_{\phi,k}(x), \sigma_{\phi,k}(x))$, generating component-specific latent representations. These representations are combined using component weights determined by our novel multivariate sigmoid activation (described in section 4.1.4), which replaces the standard sigmoid to explicitly model label dependencies. The prior distribution is formulated as a mixture of five Gaussians $p(z) = \sum_{k=1}^{5} \pi_k \mathcal{N}(z|\mu_k, \Sigma_k)$,

where each component specializes in capturing features relevant to a specific cardiac condition. Unlike conventional approaches that rely solely on latent space proximity to implicitly model label correlations, our method introduces a learnable interaction matrix that directly quantifies pairwise relationships between conditions. This matrix transforms independent component probabilities into a correlated distribution through the multivariate sigmoid function, enabling the model to capture clinically meaningful dependencies such as comorbidity patterns and mutually exclusive diagnoses while maintaining the probabilistic framework necessary for uncertainty quantification and out-of-distribution detection.

4.1.3 Network Architecture

The architecture of the proposed is shown in Fig. 4.1. It includes an encoder and a decoder network, both of which are based on 1-D adaptation of ConvNeXt_base [126] architecture. The encoder pathway consists of an initial stem layer comprising a 1D convolution operation with kernel size (k = 4) and stride (s = 4), followed by layer normalization operation that normalizes the features across the channel dimension for each temporal position independently. Following the stem layer, the encoder's main body comprises four sequential stages, each containing a series of ConvNeXt blocks. Each stage begins with a downsampling layer consisting of Layer Normalization followed by a 1D convolutional layer ($k = 2$, $s = 2$), which effectively halves the temporal resolution while increasing the channel dimensions progressively: $h_i = \text{Conv}_{2\times1}(\text{LN}(h_{i-1}))$, where h_i represents the hidden feature representation at stage i and channel dimensions expand as (128, 256, 512, 1024). Each ConvNeXt block first applies a depthwise 1D convolution ($k = 7$, $p = 3$) for local feature extraction, transforming each channel independently while maintaining spatial relationships followed by Layer Normalization. Following normalization, the features are processed through two linear layers implementing an inverted bottleneck structure, first expanding channels by 4x: $z_{\text{exp}} = W_1 \cdot z_{\text{norm}}$ where $W_1 \in \mathbb{R}^{4C\times C}$, applying GELU activation [127], then projecting back: $z_{\text{proj}} = W_2(\text{GELU}(z_{\text{exp}}))$ where $W_2 \in \mathbb{R}^{C\times 4C}$. The transformed features are scaled by a learnable parameter γ (initialized to 1e-6) and added to the input through a residual connection: $h_{\text{out}} = h_{\text{in}} + \gamma \odot z_{\text{proj}}$. The complete block transformation can be expressed as: $h_{\text{out}} = h_{\text{in}} + \gamma \odot W_2(\text{GELU}(W_1(\text{LN}(\text{DWConv}_{7\times1}(h_{\text{in}})))))$.

To enable probabilistic modeling through the GMM-VAE framework, the extracted features undergo Global Average Pooling (GAP) to obtain a fixed-dimensional representation: $h_{\text{gap}} = \text{GAP}(h_4)$ where h_4 is the output of the final stage. Formally, this is defined as: $h^c_{\text{gap}} = \frac{1}{L}\sum_{i=1}^{L} h_4^{c,i}$, where h^c_{gap} represents the c^{th} element of the output vector corresponding to the c^{th} channel, and $h_4^{c,i}$ denotes the feature value at temporal position i in channel c. This pooled representation is then transformed through two parallel dense layers to parameterize the approximate posterior $q_\phi(z|h_{\text{gap}})$. Specifically, one branch outputs the mean

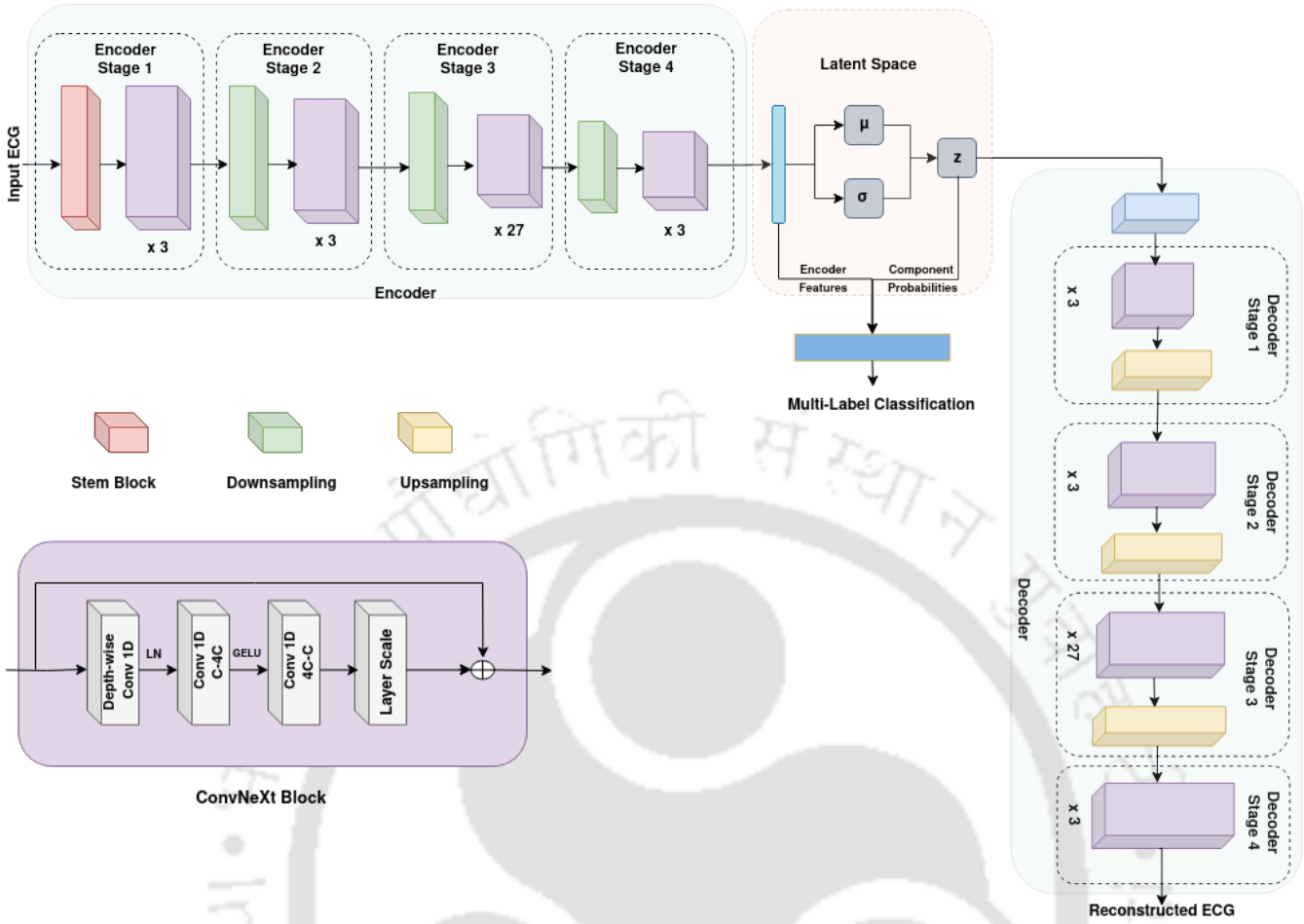

Figure 4.1: Schematic overview of the proposed approach. 1D adaptation of ConvNeXt base is used for encoder and decoder architecture.

vector $\mu_\phi(h_{\text{gap}})$ and the other outputs the log-variance vector $\log \sigma^2_\phi(h_{\text{gap}})$, enabling the sampling of latent vectors through the reparameterization trick: $z = \mu_\phi(h_{\text{gap}}) + \sigma_\phi(h_{\text{gap}}) \odot \epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$. The latent space is structured as a mixture of K Gaussian components, where K matches the number of cardiac conditions (K=5), enabling the model to capture label-specific features and their correlations.

The decoder pathway mirrors the encoder's hierarchical structure but operates in reverse, progressively recovering the temporal resolution while decreasing feature dimensions. Starting from the latent representation z, a linear projection first maps it to the initial decoder dimension of 1024 channels: $h_{\text{dec},0} = W_{\text{proj}} \cdot z$. The architecture comprises four sequential stages, with channel dimensions decreasing as $1024 \to 512 \to 256 \to 128$, mirroring the encoder's dimensions in reverse. Each stage maintains the core ConvNeXt block structure but replaces the downsampling operations with upsampling through transposed convolutions. After each stage i (except the last), a transposed convolution (kernel=2, stride=2) followed by Layer Normalization doubles the temporal resolution: $h_{\text{dec},i+1} = \text{LN}(\text{TransConv}_{2\times1}(h_{\text{dec},i}))$. The final stage

employs a transposed convolution with kernel size 4 and stride 4 to restore the original temporal resolution and channel dimensions of the input ECG signal, producing the reconstruction $\hat{x}$.

From the encoder's feature representation, we implement a specialized classification pathway for multilabel cardiac condition detection. The classification head processes the feature map h_4 (1024 channels) through a sequential structure: first applying a strided convolutional layer (kernel=3, stride=2) that reduces dimensionality to 256 channels, followed by batch normalization and GELU activation. Global context is then captured via adaptive average pooling, collapsing the temporal dimension to a single point. After flattening, a linear projection maps to the 5-dimensional output space corresponding to our target cardiac conditions. Rather than using independent sigmoid activations, we employ a correlation-aware multivariate sigmoid function that models interdependencies between cardiac conditions.

4.1.4 Multivariate Sigmoid

To model interdependencies between cardiac conditions, we implement a correlation-aware multivariate sigmoid function that extends beyond independent binary classifications. Given a vector of logits $l \in \mathbb{R}^5$ and a condition interaction matrix $M \in \mathbb{R}^{5\times5}$, we first compute marginal probabilities through the standard sigmoid function: $P_{\text{marginal}} = \sigma(l)$, where $\sigma(l_i) = \frac{1}{1+e^{-l_i}}$. The interaction matrix M is initialized as a learnable parameter and optimized during model training, allowing the network to discover and encode condition relationships directly from the data rather than relying on predefined clinical associations. This data-driven approach enables the model to capture both known and potentially novel condition interdependencies. The interaction matrix captures pairwise relationships between conditions, with M_{ij} representing the influence of condition j on condition i. We symmetrize this matrix ($M_{\text{sym}} = \frac{1}{2}(M + M^T)$) and remove self-interactions by element-wise multiplication with $(1 - I)$, where I is the identity matrix. For each condition i, the correlated probability is computed using our multivariate sigmoid function:

$$\sigma_{MV}(z_i, l)_k = \text{clamp}\left(\sigma(z_i, k) + \frac{1}{2}\tanh\left(\sum_{j\neq i} M_{k,j} \cdot \sigma(z_j, j)\right), 0, 1\right) \tag{4.1}$$

This formulation ensures that the prediction for each condition k is influenced by the marginal probabilities of other conditions, weighted by their interaction coefficients. The hyperbolic tangent function constrains the adjustment magnitude, while the clamp operation ensures valid probability values. The learned correlation matrix directly influences prediction behavior through a bidirectional adjustment mechanism. For conditions with positive correlation coefficients ($M_{i,j} > 0$), the detection of one condition automatically enhances the prediction probability of its correlated counterpart. Conversely, negative coefficients ($M_{i,j} < 0$) enable the

model to capture mutual exclusivity, where the presence of one condition reduces the likelihood of another. This dynamic adjustment operates proportionally to both the confidence level of the detected primary condition and the magnitude of the learned correlation coefficient. The tanh function in the adjustment term ensures these modifications remain bounded, preventing excessive probability shifts while still allowing meaningful adjustments when strong correlations are detected. This mechanism effectively captures both co-occurrence and exclusion patterns learned from the training data, resulting in a prediction framework that respects the underlying statistical dependencies between conditions without requiring explicit rules or thresholds.

4.1.5 Composite Training Objective

The proposed network architecture must be optimized for multiple concurrent objectives to achieve effective uncertainty-aware multi-label ECG classification while preserving the intrinsic relationships between cardiac conditions. Our training objective integrates three key components: (1) a reconstruction term that ensures the decoder accurately reproduces input ECG morphology, preserving critical diagnostic features; (2) a Kullback-Leibler divergence term that regularizes the latent space structure toward our mixture of Gaussians prior, encouraging component specialization; and (3) a classification loss that leverages the correlation-aware multivariate sigmoid to improve multi-label prediction accuracy. For ECG reconstruction, we adopted Huber loss, which combines the characteristics of Mean Squared Error and Mean Absolute Error. It behaves quadratically for small errors and linearly for large ones, with a transition controlled by hyperparameter β, making it robust to outliers while remaining sensitive to subtle ECG morphology changes. Given a pair of reconstructed ECG $\hat{x}$ and the target x, we compute the reconstruction loss as:

$$\mathcal{L}_{Rec} = \begin{cases} 0.5(\hat{x}_k - x_k)^2, & \text{if } |\hat{x}_k - x_k| < \beta. \\ \beta(|\hat{x}_k - x_k| - 0.5\beta), & \text{otherwise} \end{cases} \tag{4.2}$$

To regularize the latent space to follow the GMM prior, where the prior distribution $p_\theta(z)$ is defined as a mixture of Gaussian components equal to the number of labels: $p_\theta(z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(z|\mu_k, \Sigma_k)$, we employ the KL divergence loss, it measures the discrepancy between the encoder's output distribution $q_\phi(z|x)$ and the GMM prior $p_\theta(z)$, where $p_\theta(z) = \sum_{k=1}^{K} \pi_k \mathcal{N}(z|\mu_k, \Sigma_k)$ and is given as :

$$\mathcal{L}_{\text{KL}} = D_{\text{KL}}(q_\phi(z|x)|p_\theta(z)) \tag{4.3}$$

This formulation ensures that the latent representations follow a mixture of Gaussians with K components,

each potentially corresponding to a distinct pattern of label co-occurrence. The KL divergence term can be expanded as: $\mathbb{E}q\phi(z|x)[\log q_\phi(z|x) - \log(\sum_{k=1}^{K} \pi_k \mathcal{N}(z|\mu_k, \Sigma_k))]$, which encourages the encoder to produce latent codes that are well-structured and aligned with the GMM components.

The classification loss leverages a dedicated convolutional classification head that processes encoder features to predict cardiac conditions through our multivariate sigmoid activation to generate correlation-aware multi-label predictions. The logits $l_i \in \mathbb{R}^5$ from the classification head are transformed through our multivariate sigmoid function to obtain correlation-aware label probabilities, $\hat{y}_i = \sigma_{\text{MV}}(l_i, M)$. Where M is the learnable interaction matrix capturing dependencies between cardiac conditions, and σ_{MV} is the multivariate sigmoid. The binary cross-entropy loss for multi-label classification is then formulated as:

$$\mathcal{L}_{\text{class}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}[y_{i,k}\log(\hat{y}_{i,k}) + (1 - y_{i,k})\log(1 - \hat{y}_{i,k})] \tag{4.4}$$

Where N is the number of samples, K is the number of labels, $y_{i,k}$ is the ground truth (0 or 1) for label k in sample i and $\hat{y}_{i,k}$ is the predicted probability after applying the multivariate sigmoid.

These individual loss components are combined using homoscedastic uncertainty-based loss balancing [128], which automatically learns the relative weights of each loss term during training. The total loss is formulated as:

$$\mathcal{L}_{\text{total}} = \frac{1}{2\sigma_1^2}\mathcal{L}_{\text{Rec}} + \frac{1}{2\sigma_2^2}\mathcal{L}_{\text{KL}} + \frac{1}{2\sigma_3^2}\mathcal{L}_{\text{class}} + \log(\sigma_1\sigma_2\sigma_3) \tag{4.5}$$

where σ_1, σ_2, and σ_3 are learnable parameters representing the uncertainty associated with each task. These parameters are optimized during training, automatically adjusting the relative importance of each loss component based on their inherent uncertainties. The weighting terms $\frac{1}{2\sigma_i^2}$ ensure positive weights and provide numerical stability, while the regularization term $\log(\sigma_1\sigma_2\sigma_3)$ prevents the uncertainties from becoming too large and acts as a soft constraint on their magnitude, ensuring training stability.

4.1.6 OOD detection

OOD detection refers to identifying test samples that differ significantly from the training distribution. Formally, given a model trained on data from a distribution $P_{in}(x)$, an OOD detection system aims to determine whether a test sample x is drawn from $P_{in}(x)$ or from a different distribution $P_{out}(x)$. The ability to identify unfamiliar patterns is particularly important in medical applications where encountering novel patterns or pathologies not seen during training is common and could lead to unreliable predictions. The OOD detection task can be formulated as learning a decision function g(x) that discriminates between in-distribution and OOD samples:

$$g(x) = \begin{cases} 1, & \text{if } x \sim P_{in}(x) \text{ (ID)} \\ 0, & \text{if } x \sim P_{out}(x) \text{ (OOD)} \end{cases} \tag{4.6}$$

Within our multi-label GMM-VAE framework, we investigate two complementary approaches for implementing this decision function. The first approach utilizes Reconstruction Error, which quantifies the model's reconstruction fidelity. This metric leverages the principle that the VAE is optimized to minimize reconstruction loss on the training distribution $P_{\text{in}}(x)$, resulting in higher error values for samples from $P_{\text{out}}(x)$. The reconstruction error is computed as:

$$S_{\text{RE}}(x) = ||x - \hat{x}||_2^2 \tag{4.7}$$

where $\hat{x}$ is the reconstructed output of the VAE for input x.

The second approach employs Average Component Entropy (ACE), capturing the uncertainty in component assignment within the latent GMM space:

$$S_{\text{ACE}}(x) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \gamma_{i,k} \log(\gamma_{i,k}) \tag{4.8}$$

where $\gamma_{i,k}$ represents the responsibility (posterior probability) of component k for sample i. Higher entropy indicates ambiguous component assignment, which is expected for samples drawn from $P_{\text{out}}(x)$.

A decision function is implemented using a threshold τ on these scoring functions:

$$g(x) = \begin{cases} 1, & \text{if } S(x) \leq \tau \\ 0, & \text{if } S(x) > \tau \end{cases} \tag{4.9}$$

The threshold is determined using conformal prediction to provide statistical guarantees:

$$\tau = \text{Quantile}_{(n+1)(1-\alpha)/n}(\{S(x_i)\}_{i=1}^{n}) \tag{4.10}$$

with $\alpha = 0.05$, ensuring that approximately 95% of in-distribution samples will be correctly classified.

4.2 Experimental Setup

In this section, we detail the experimental approach used to validate our proposed framework. We begin by introducing the datasets, comprising both in-distribution and out-of-distribution clinical ECG databases. Next, we describe the preprocessing steps applied to the ECG signals, followed by details regarding

comparative analysis of different approaches. We then outline the evaluation metrics used to assess various aspects of the task within our framework. Finally, we provide a comprehensive overview of the framework's implementation, including network architecture, training procedures, and hyperparameter configurations.

4.2.1 Datasets

In-distribution: Similar to previous chapters, we employ the *PTB-XL* dataset [91] for model development and evaluation. In this work, we leverage the five broad diagnostic superclasses available in the dataset, representing co-occurring pathologies: MI, CD, STTC, HYP, and NORM. Each ECG recording may be associated with multiple labels. We adhere to the recommended stratified split (train: fold 1-8, val: fold 9 and test: fold 10), which ensures consistent distribution of diagnoses and demographic factors across subsets while maintaining patient-wise separation to prevent data leakage. This standardized partitioning enables fair comparison with previous approaches and ensures the robustness of our evaluation metrics.
Out-of-distribution: We incorporate two external datasets: the China Physiological Signal Challenge (CPSC) [108, 109] dataset and the Georgia 12-lead ECG Challenge (G12EC) dataset [109] for evaluating OOD performance. The CPSC dataset contains 6,877 12-lead ECGs from Chinese patients with nine diagnostic classes, while the G12EC dataset includes 10,344 ECGs from Emory University Hospital with diverse cardiac conditions. These datasets feature different demographics, equipment and protocols compared to PTB-XL, making them ideal for assessing generalization to unfamiliar patterns and novel cardiac conditions. Specifically, we consider diagnostic labels from these datasets that are not covered within the PTB-XL dataset's superclasses, treating them as OOD samples. This includes rhythmic conditions such as atrial fibrillation, atrial flutter, and various morphological forms that represent distinct electrophysiological patterns not explicitly represented in our training data.

4.2.2 Preprocessing

We follow the same preprocessing protocol established in previous chapters. First, we apply a moving average filter [110] to eliminate baseline wander, followed by a nonlocal means (NLM) filter [94] to suppress muscle artifacts and high-frequency noise. The filtered signals undergo Z-score normalization, with quality assessment based on established criteria [95]. For consistency across databases, all signals are resampled to a uniform 250 Hz sampling rate, and 10-second segments are extracted from each record to maintain consistent input length. Our experimental protocol uses PTB-XL folds 1-8 for training, fold 9 for validation, and the test fold for multi-label classification evaluation, while reserving the G12EC and CPSC datasets

specifically for out-of-distribution detection assessment.

4.2.3 Comparative Analysis

To empirically validate the efficacy of the proposed method, we conducted multiple groups of comparative experiments. The details of the experiments are as follows.

- **Ablation Study**:

  The effectiveness of the proposed GMM-VAE architecture for multi-label ECG classification is validated through a series of systematic ablation studies: We compare our full model against variants that remove key components: (1) replacing the GMM-VAE with a standard VAE to assess the importance of mixture modeling, (2) using the encoder features directly, and (3) using sigmoid activation directly inplace of the multivariate sigmoid.

- **Quantitative Comparison**: We validate our approach through comprehensive comparison with leading methods in multi-label ECG classification, including both foundational deep learning architectures (ResNet [103], DenseNet [129], ConvNeXt [126]) and state-of-the-art specialized frameworks (ASTLNet [44], Conv-RGNN [130]). All models were implemented following the same preprocessing and training protocols to ensure fair comparison.

4.2.4 Evaluation Methods

We evaluate the proposed approach across three key areas: classification performance, reconstruction performance, and OOD detection. For classification performance evaluation, we employ standard multi-label classification metrics: the Hamming Loss, which measures the fraction of incorrectly predicted labels as $\mathcal{H}_{\text{loss}} = \frac{1}{NK}\sum_{i=1}^{N}\sum_{k=1}^{K}|y_{i,k} - \hat{y}_{i,k}|$, where N is the number of samples and K is the number of labels; F1 Score, which computes the harmonic mean of precision and recall for each instance as $F1 = \frac{1}{N}\sum_{i=1}^{N}\frac{2|\hat{Y}_i \cap Y_i|}{|\hat{Y}_i|+|Y_i|}$, where $\hat{Y}_i$ and Y_i represent the predicted and true label sets for instance i; Coverage, which measures how far on average we need to go down the ranked list of labels to cover all the relevant labels of an instance, defined as $\text{Coverage} = \frac{1}{N}\sum_{i=1}^{N}\max_{k:y_{i,k}=1}\text{rank}(y_{i,k}) - 1$, where $\text{rank}(y_{i,k})$ is the position of label k in the prediction ranking for instance i; AUPRC, which summarizes the precision-recall trade-off across different threshold values for each label and averages across all labels as $\text{AUPRC} = \frac{1}{K}\sum_{k=1}^{K}\int_0^1 \text{precision}_k(r)dr$, where $\text{precision}_k(r)$ is the precision at recall level r for label k; and AUROC, which quantifies the model's ability to discriminate between positive and negative classes across all possible thresholds.

For evaluating reconstruction performance, we employ the same metrics used in Chapter 3 to ensure consistent assessment: Percentage Root-Mean-Square Difference (PRD), which quantifies the normalized error between original and reconstructed signals; Structural Similarity Index (SSIM) [115], which measures the preservation of structural information in the signal; Wavelet Energy based Diagnostic Distortion (WEDD) [114], which evaluates distortion in clinically relevant frequency bands; and Dynamic Time Warping (DTW), which measures similarity between temporal sequences by calculating the optimal alignment between the original and reconstructed signals. For OOD detection capability, we implement two complementary approaches: reconstruction error analysis and ACE computed from GMM component responsibilities. We evaluated both methods on external OOD datasets, using standard metrics including AUROC and AUPRC. The test sets contained a mixture of in-distribution samples and OOD samples, allowing to assess detection performance under practical conditions. For clinical relevance, we also report condition-specific metrics including accuracy (ACC), sensitivity (Sen), specificity (Spe), and F1 Score value for each cardiac condition.

4.2.5 Implementation Details

The implementation was carried out in Python using PyTorch (v2.0.1) and executed on a system with an an Intel(R) Xeon(R) Silver 4214R 32-core CPU @ 2.40 GHz, 64 GB of RAM, and an NVIDIA A100-PCIE GPU, with development carried out in Visual Studio Code. Training employed the AdamW [131] optimizer with a starting learning rate of 0.001 and a mini-batch size of 32. A learning rate scheduler was used to reduce the learning rate by a factor of 5 if validation loss did not improve over 10 epochs. To further enhance generalization, early stopping was employed, halting training if no improvement was observed for 25 epochs. The model corresponding to the lowest validation loss was selected for downstream analysis.

4.3 Results and Discussions

In this section, we present a comprehensive evaluation of our proposed GMM-VAE framework across three key aspects: multi-label classification performance, ECG reconstruction quality, and out-of-distribution detection. We then provide a comparative analysis featuring both an ablation study and quantitative comparisons with existing approaches.

4.3.1 Performance on reconstruction quality

We evaluate various aspects of reconstruction quality using four distinct metrics described in the section 4.2.4. Table 4.1 summarizes the reconstruction results across all 12 ECG leads. The overall results

demonstrate high-fidelity reconstruction across all the leads. The PRD, which measures the normalized error between original and reconstructed signals, shows consistent performance across all signals with values ranging from 0.2961 to 0.4254, achieving a mean PRD of 0.3403 ± 0.0390. This relatively low PRD value and small standard deviation indicate stable reconstruction quality across different ECG morphologies. The WEDD, quantifies the preservation of diagnostic information through wavelet decomposition, demonstrates strong performance with values ranging from 0.2843 to 0.4060, yielding a mean of 0.3259 ± 0.0368. The lower WEDD scores indicate minimal distortion of clinically relevant ECG components across different frequency bands, suggesting that the model effectively preserves diagnostic features in the reconstructed signals. The DTW analysis reveals effective temporal alignment across most ECG leads, with an overall mean of 9.7221±1.3451, though certain leads like lead 3 (12.7680) exhibit greater reconstruction challenges than better-performing leads like lead aVL (8.3945). The SSIM, which evaluates the perceptual quality of reconstruction, shows strong performance with values ranging from 0.6432 to 0.7839, achieving a mean of 0.7240 ± 0.0432. The high SSIM scores, particularly above 0.70 in eight out of twelve leads, indicate excellent preservation of structural ECG characteristics. Notably, lead V4, lead V5, and lead V6 achieved the highest SSIM values (0.7838, 0.7839, and 0.7583 respectively), suggesting particularly strong reconstruction of these specific ECG morphologies. Figure 4.2 presents a comprehensive comparison of original and reconstructed signals across all twelve leads of a representative ECG record. The close alignment between original and reconstructed signals across all leads corroborates our quantitative findings and demonstrates the model's capability to maintain diagnostic integrity across the complete spatial representation of cardiac electrical activity. The reconstruction quality remains consistent across leads with different signal amplitudes and morphological complexities, suggesting robust performance across varying ECG characteristics.

Table 4.1: ECG Reconstruction Performance across PTB-XL Test Set

	Leads												Mean ± Std
Metrics	I	II	III	aVL	aVR	aVF	V1	V2	V3	V4	V5	V6	
PRD ↓	0.3138	0.3325	0.4254	0.2984	0.3834	0.3959	0.3459	0.3336	0.3306	0.3168	0.2961	0.3105	0.3403 ± 0.0390
DTW ↓	9.0598	9.3454	12.7680	8.3945	11.4925	11.5306	9.6168	8.8248	9.0843	9.1825	8.5960	8.7698	9.7221 ± 1.3451
WEDD ↓	0.3022	0.3189	0.4060	0.2877	0.3675	0.3788	0.3323	0.3180	0.3135	0.3011	0.2843	0.3003	0.3259 ± 0.0368
SSIM ↑	0.7188	0.7132	0.6432	0.7346	0.6809	0.6788	0.6860	0.7420	0.7645	0.7838	0.7839	0.7583	0.7240 ± 0.0432

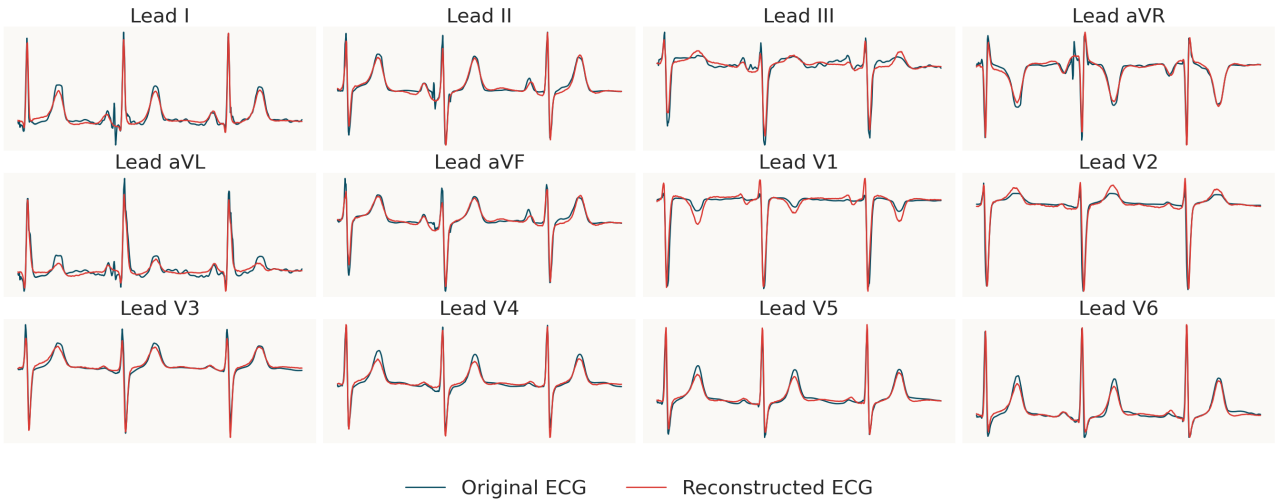

Figure 4.2: Comparison between the reconstructed and original signals.

4.3.2 Performance on OOD Detection

Table 4.2 gives the result of OOD detection of CPSC and G12EC dataset. On CPSC dataset, the reconstruction error approach achieved an AUROC of 0.8236 and AUPR of 0.8004, indicating good separation between in-distribution and OOD samples. However, the ACE method demonstrated superior performance with an AUROC of 0.8675 and AUPR of 0.8412, representing a 4.4% improvement in AUROC and 4.1% in AUPR. The performance gap was considerably more pronounced on G12EC, which proved more challenging for both methods. While reconstruction error struggled with an AUROC of only 0.6878 (though maintaining a reasonable AUPR of 0.8477), the ACE method significantly outperformed it with an AUROC of 0.7950 and AUPR of 0.8689, a substantial 10.7% improvement in AUROC. These results suggest that Average Component Entropy is more robust to dataset variations and provides more reliable OOD detection, particularly on challenging datasets. The ACE method's stronger performance can be attributed to its ability to capture uncertainty in component assignment within the GMM latent space, which appears to be a more sensitive indicator of distribution shift than simple reconstruction fidelity. This advantage is particularly evident when facing more complex OOD detection scenarios, as demonstrated by the results on G12EC dataset.

Method	CPSC		G12EC	
	AUROC	AUPRC	AUROC	AUPRC
Reconstruction Error	0.8236	0.8004	0.6878	0.8477
Average Component Entropy	0.8675	0.8412	0.7950	0.8689

Table 4.2: Performance on OOD for CPSC and G12EC datasets

4.3.3 Performance on multi-label classification

The performance of our proposed framework in multi-label classification is evaluated through six complementary metrics that assess different aspects of the classification quality. Table 4.3 presents the multi-label classification performance of the proposed network on the PTB-XL dataset (Test-Fold).The model achieves a Hamming Loss of 0.1344, indicating that only 13.44% of label predictions are misclassified across all instance-label pairs. The Coverage score of 1.7379 reveals that on average, we need to examine fewer than two positions in the ranked prediction list to capture all relevant labels for each ECG sample. The RLoss value of 0.1076 indicates that the model is effective in ranking relevant labels higher than irrelevant ones, with only about 10.8% of label pairs being incorrectly ranked. Most notably, the model demonstrates excellent discriminative ability with an F1 Score of 74.86, suggesting effective balancing of precision and recall across all five classes. The high AUROC of 90.64 reflects robust threshold-independent discrimination, while the PRAUC of 79.25 highlights strong performance in the precision–recall space, which is particularly important in this clinical setting characterized by class imbalance.

Table 4.3: Multi-label classification performance of the proposed network.

Database	HLoss ↓	Coverage ↓	RLoss ↓	F1 Score ↑	AUROC ↑	PRAUC ↑
PTB-XL (Test-Fold)	0.1344	1.7379	0.1076	74.86	90.64	79.25

The label-wise evaluation of our model on the PTB-XL test set reveals varying performance across different cardiac conditions, with particularly strong results for normal ECG patterns, CD and ST/T changes. The model shows an accuracy in identifying STTC with 87.21% accuracy and a high AUROC of 91.48%, demonstrating robust discrimination capability for these subtle ECG variations. Normal (NORM) ECG patterns achieve a superior balanced performance with an accuracy of 86.35%, F1 score of 85.19%, and excellent sensitivity-specificity trade-off (89.61%, 83.8%). For MI, the model maintains good overall performance with 85.07% accuracy and an AUROC of 89.66%, showing particularly strong specificity (87.55%) in detecting this critical condition. CD demonstrates the highest specificity (92.00%) among all conditions with good accuracy (87.67%) and F1 score (72.76%). This suggests the model excels at identifying the characteristic QRS complex widening and morphological alterations associated with conduction abnormalities. HYP presents the most challenging classification task, with lower but still acceptable performance metrics (accuracy: 84.63%, F1 score: 44.97%, AUROC: 80.15%), suggesting room for improvement in detecting this particular condition. Notably, the model maintains high specificity

across all conditions (ranging from 83.80% to 92.00%), indicating reliable performance in minimizing false positives, a crucial aspect for clinical applications.

Table 4.4: Label-wise performance of multi-label cardiac ailments classification on PTB XL test set.

Database	label	Acc	F1 Score	Spe.	Sen.	AUROC
PTB-XL (Test-Fold)	NORM	86.35	85.19	83.80	89.61	92.23
	MI	85.07	72.29	87.55	77.67	89.66
	HYP	84.63	44.97	89.66	51.14	80.15
	CD	87.67	72.76	92.00	72.83	90.04
	STTC	87.21	74.43	89.92	78.50	91.48

4.3.4 Ablative study

We conduct comprehensive ablation studies to validate the effectiveness of various components in our proposed framework. In evaluating architectural components, we first analyze the impact of the GMM-VAE structure by comparing it with a standard VAE baseline, observing that the GMM-VAE achieves superior performance with an improvement of 0.65% in F1 score (73.18 vs. 72.53) and 1.79% in AUROC (89.30 vs. 87.51), demonstrating the effectiveness of mixture modeling in capturing complex label dependencies. The introduction of our correlation-aware sigmoid function provides additional gains, with our complete framework achieving an F1 score of 74.86 and AUROC of 90.64, representing improvements of 1.68% and 1.34% respectively over the standard sigmoid approach. The *GMM-VAE with sigmoid* approach, which incorporates Gaussian mixture modeling but lacks our correlation-aware formulation, ranks second with notable but lower performance metrics across all evaluation criteria. The *Standard VAE* baseline follows with moderate performance (F1: 72.53, AUROC: 87.51), while the *Direct Encoder* approach shows the weakest results overall (F1: 72.10, AUROC: 87.18), highlighting the benefits of probabilistic modeling. These findings validate our architectural design choices, particularly emphasizing the significant improvement gained by incorporating our multivariate sigmoid with learned correlation modeling compared to standard sigmoid approaches. The consistent enhancement across all metrics, including a reduction in Hamming Loss to 0.1344 from 0.1484 in the next best method, demonstrates that our framework effectively captures the complex relationships between cardiac conditions, resulting in more accurate and reliable multi-label classification that better reflects the interdependencies present in real-world cardiac diagnostics.

Table 4.5: Ablative study on Multi-label classification performance.

Method	HLoss ↓	Coverage ↓	RLoss ↓	F1 Score ↑	AUROC ↑	PRAUC ↑
Standard VAE	0.1356	1.7283	0.1092	72.53	87.51	72.08
Direct Encoder	0.1491	1.7456	0.1132	72.10	87.18	71.35
GMM-VAE with sigmoid	0.1484	**1.7306**	0.1098	73.18	89.30	74.24
This Work	**0.1344**	**1.7379**	**0.1076**	**74.86**	**90.64**	**79.25**

4.3.5 Quantitative Comparison

We compare our approach with several state-of-the-art methods for multi-label ECG classification. The baseline methods include: standard deep learning approaches like ResNet, DenseNet and ConvNeXt adapted for multi-label classification and recent specialized approaches (ASTLNet, Conv-RGNN). Among the foundational deep learning architectures, ResNet [103] shows modest performance (F1: 71.30, AUROC: 87.26), while DenseNet [129] achieves slightly better results with an F1 score of 71.81 and AUROC of 87.88. ConvNeXt [126], a more recent architecture, demonstrates incremental improvements (F1: 72.10, AUROC: 87.18) over these traditional networks. Moving to specialized ECG analysis frameworks, ASTLNet [44] shows substantial performance gains with an F1 score of 73.58 and the second-highest AUROC (91.31), reflecting the benefits of its attention-based temporal modeling approach. Conv-RGNN [130] achieves strong results across all metrics (F1: 74.42, AUROC: 90.14, PRAUC: 78.37), leveraging its graph-based representation of inter-lead relationships. Our proposed method achieves the best performance across all evaluation criteria, with the lowest error rates (HLoss: 0.1344, RLoss: 0.1076) and highest discriminative metrics (F1: 74.86, AUROC: 90.64, PRAUC: 79.25). The performance improvements are particularly notable in PRAUC (0.88% improvement over Conv-RGNN), indicating enhanced precision-recall balance, which is critical in clinical applications with class imbalance. While the margins of improvement vary across metrics, the consistent superiority across all six evaluation criteria validates our integrated approach of structured latent space modeling with correlation-aware classification. Beyond improved accuracy, our framework provides an additional capability that these baseline methods lack: principled uncertainty quantification through the GMM latent space, enabling the model to abstain from predictions on ambiguous or out-of-distribution samples. This uncertainty-aware diagnosis comes with minimal computational overhead, achieving an inference time of 19.76 ± 0.36 ms per sample, which is comparable to the baseline methods while providing both classification and uncertainty estimates in a single forward pass. These results demonstrate that explicitly modeling both feature-level representations and label-level dependencies, combined with uncertainty quantification, yields meaningful performance gains and enhanced clinical safety

in multi-label ECG diagnostic tasks.

Table 4.6: Quantitative Comparison with state-of-the-art approaches

Method	HLoss ↓	Coverage ↓	RLoss ↓	F1 Score ↑	AUROC ↑	PRAUC ↑
ResNet [103]	0.1401	1.7529	0.1209	71.30	87.26	71.76
DenseNet [129]	**0.1345**	1.7515	0.1181	71.81	87.88	72.79
ConvNeXt [126]	0.1415	1.7456	0.1132	72.10	87.18	71.35
ASTLNet [44]	0.1162	**1.6892**	**0.1009**	73.58	**91.31**	77.91
Conv-RGNN [130]	0.1347	1.7575	0.1183	74.42	90.14	78.37
This Work	**0.1344**	1.7379	0.1076	**74.86**	90.64	**79.25**

4.4 Summary

This chapter presents a novel probabilistic framework for multi-label ECG classification that simultaneously addresses three critical challenges in automated cardiac diagnosis: accurate multi-label classification, uncertainty quantification, and out-of-distribution detection. The proposed approach integrates a specialized ConvNeXt-based architecture with a Gaussian Mixture Model Variational Autoencoder and a correlation-aware sigmoid function with a learnable interaction matrix that captures interdependencies between cardiac conditions. We employed ECG reconstruction as an auxiliary task, which contributes to learning the underlying distribution of ECG signals for richer latent representations while simultaneously enhancing the primary classification performance and delivering high-fidelity signal reconstruction across all 12 leads (mean SSIM: 0.7240 ± 0.0432, mean PRD: 0.3403 ± 0.0390). Experiments utilizing the PTB-XL dataset demonstrated that our approach achieved superior classification performance across all metrics (HLoss: 0.1344, F1 Score: 74.86) compared to five state-of-the-art baseline methods, with particularly high metrics for normal ECGs (F1: 85.19, AUROC: 92.23) and ST/T changes (AUROC: 91.48). Ablation studies confirmed the significant contributions of both the GMM-VAE structure and the correlation-aware sigmoid, with each component providing measurable performance improvements. External validation using the CPSC and G12EC datasets verified the framework's ability to recognize unfamiliar cardiac patterns through complementary OOD detection approaches based on reconstruction error analysis and Average Component Entropy. The comprehensive evaluation validates our integrated approach, which advances automated ECG analysis by addressing the inherent complexity of cardiac diagnosis, moving beyond simple classification to deliver more clinically relevant outputs while providing essential safety features for potential clinical deployment through effective uncertainty quantification and OOD detection capabilities.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

5

Detecting Rhythmic and Morphological Abnormalities from ECG plot Images

Contents

In the previous chapters, we explored approaches for knowledge-aware automated cardiovascular diagnosis using digital ECG signals. However, a significant portion of clinical ECG data exists in the form of paper-style plots, including both physical printouts and digital image formats (such as PDFs) that replicate the traditional paper layout, particularly in resource-constrained healthcare settings and for accessing historical patient records. Healthcare facilities typically employ short-duration 12-lead ECG setup for cardiac assessment, where multi-lead signals are collected and stored as standardized paper plot [132, 133]. These ECG plots present both rhythm strips and standard 12-lead recordings in a structured layout, displaying different views of the heart through distinct lead tracings. Fig. 5.1 shows the standard ECG paper representation (henceforth referred to as ECG images). Manual interpretation of these ECG images is challenging even for experienced cardiologists, requiring systematic analysis of different regions, examining the rhythm strip for cardiac rhythm and the 12-lead layout for morphological abnormalities. This process of visual inspection is not only tedious but also susceptible to subjective variations, highlighting the need for automated interpretation systems. However, adapting signal-based models for ECG image analysis presents unique challenges due to the fundamental difference in data representation: while digital ECG signals treat each lead as a separate 1D signal, ECG images represent the 12 leads in a 2D spatio-temporal arrangement.

Recent attempts at automated ECG image interpretation [76, 77, 134], while showing promising results, face several limitations. Most approaches focus on either rhythmic or morphological abnormalities, contrasting with clinical practice where both aspects are analyzed simultaneously. Furthermore, existing methods generally treat the overall ECG image representation as a uniform entity, overlooking the structured layout where specific regions serve distinct diagnostic purposes. The standard ECG layout includes specific regions for the analysis of rhythmic or morphological abnormalities, as illustrated in Fig. 5.1. Additionally, while multiple cardiac disorders commonly co-exist in ECG recordings, most current approaches treat classification as a single-label problem. This oversimplification fails to capture the complex nature of cardiac pathologies, where conditions like MI can co-occur with rhythm disturbances, conduction abnormalities, or other structural changes.

In this chapter, we address these challenges by introducing a novel approach to ECG image analysis that leverages both morphological and rhythmic information simultaneously. Reflecting upon the diagnostic approach followed by clinicians, where specific regions are precisely examined for different sets of abnormalities, we have designed a novel end-to-end multi-task framework for the detection of cardiac abnormalities through ECG images. In this framework, we modelled the classification of morphological ailments and rhythmic abnormalities as two separate tasks. Also, given the variation in regions of spatial-temporal alignment of the 12 leads and the rhythm strip in the standard representation of the ECG image,

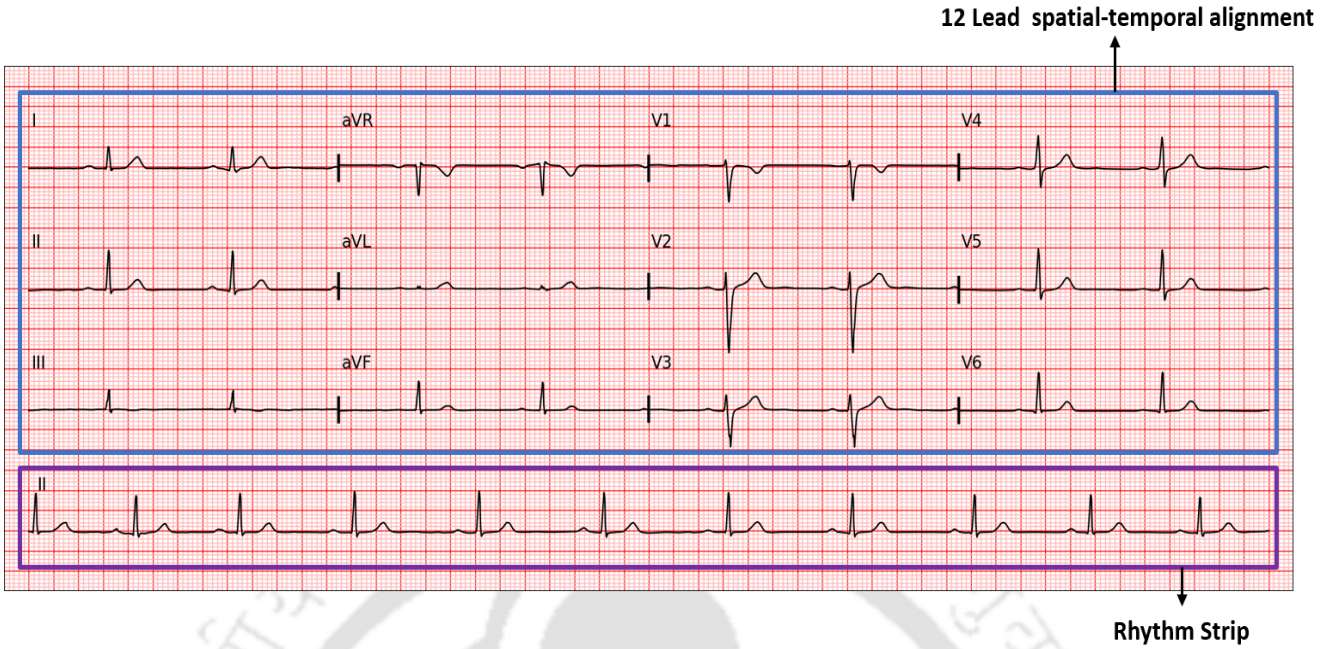

Figure 5.1: an illustration of standard representation of ECG paper tracing with first three rows representing the spatial-temporal alignment of the 12 leads and the last row representing the rhythm strip.

we introduced an additional auxiliary task of delineating the corresponding regions. This setup allows the network to focus on the 12-lead arrangement for learning morphological features and the rhythm information from the continuous recordings below the 12-lead arrangement. We integrated homoscedastic uncertainty-based dynamic weighting strategy, consistent with our approach in chapter 3, to balance the overall multi-task loss. The weight assigned to each task is inversely proportional to the level of uncertainty inherent in that task. This ensures that the model is penalized more for errors made in tasks with higher uncertainty, emphasizing the importance of reducing uncertainties in those tasks.

The remainder of this chapter is organized as follows. Section 5.1 presents our proposed multi-task learning approach for ECG image analysis. The experimental setup and implementation details are described in Section 5.2. Section 5.3 presents comprehensive results and discussion of our analysis. Finally, Section 5.4 concludes the chapter with a summary of our findings and their implications for automated ECG image interpretation.

5.1 Proposed Method: Multi-Task Learning for ECG Images

The development of this framework is motivated by the observation that the standard ECG image representation provides separate regions for assessing morphological abnormalities and rhythmic variations. To effectively utilize this region-specific information, we formulate cardiovascular abnormalities classification

into a multi-task learning framework. This approach encompasses two primary tasks: morphological cardiac ailment classification and cardiac rhythm classification, alongside an auxiliary task that delineates the corresponding regions. For each ECG image input, our methodology consists of three sequential stages: first, a backbone network extracts shared features relevant to all tasks; second, we apply region-based pooling to separate these features into distinct spatial representations corresponding to the rhythm strip and 12-lead segments; finally, these specifically pooled features are directed to task-specific classification networks. Figure. 5.2 provides a schematic overview of this proposed approach.

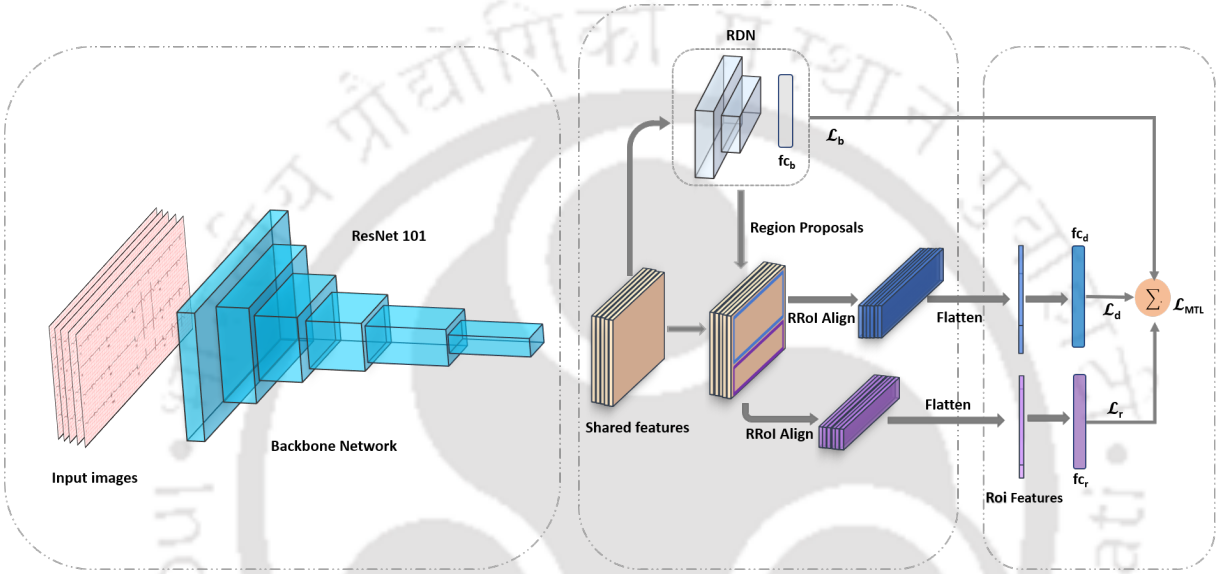

Figure 5.2: Schematic overview of the proposed approach. ResNet101 is used as backbone network to extract the shared features. The task-specific sub-networks fc_d and fc_r are jointly trained by minimizing the composite multi-task learning loss $\mathcal{L}_{MTL}$.

5.1.1 Problem definition

This work addresses two primary objectives: jointly classifying morphological cardiac ailments and rhythm patterns, alongside an auxiliary task of delineating the respective regions within ECG images. We formulate morphological ailment classification (Task 1) as a multi-label problem where each instance may simultaneously exhibit multiple conditions, while rhythm classification (Task 2) is structured as a multi-class classification problem. The region delineation task is approached as a regression problem. Let there be N instances in dataset $\mathcal{D}$=$\{(x_n, y_n^d, y_n^r, y_n^b)|1 \leq n \leq N\}$ with U disease classes and V rhythm classes. In this context, x_n represents an instance, while y_n^d, y_n^r, y_n^b correspond to the respective ground truths for cardiac ailments, cardiac rhythm, and region-specific bounding boxes. Consider $\mathcal{X} \in \mathbb{R}^D$ be input space with dimension D and $\mathcal{Y}$ denote the label space, where $\mathcal{Y}$ constitutes three parts: $\mathcal{Y}^r$ for rhythm classification task, $\mathcal{Y}^d$ for ailments classification task and $\mathcal{Y}^b$ for region delineation task. As such, the objective is to learn

a hypothesis per task as $f_t(x;\theta_s,\theta_t) : \mathcal{X} \rightarrow \mathcal{Y}$. where θ_s represents the shared parameters and θ_t refers to task-specific parameters for $t \in T$ and T represents the total number of task ($T = 3$).

An instance x_n in input space $\mathcal{X}$ is associated with labels $y_n^d \in \{0,1\}^U$, $y_n^r \in \mathcal{R}^V$, and $y_n^b \in \mathcal{R}^{10}$. Let $\mathcal{Y}^d = \{l_1, l_2, ..., l_u\} \in \{0,1\}^U$ be a finite space of U class labels, where a tuple $l_u = 1$ if u^{th} class is associated with the input instance x_n and $l_u = 0$ otherwise. For the multi-label diagnostic classification task, the aim is to learn a hypothesis $f^d : \mathcal{X} \rightarrow 2^{\mathcal{Y}^d}$ from the training set $\mathcal{D}$. In case of multiclass rhythm classification, the objective hypothesis $f^r : \mathcal{X} \rightarrow \mathcal{Y}^r$ with $\mathcal{Y}^r \in \{1, 2, ..., V\}$. For the region delineation task, we define a quintuple $\mathcal{B} \in (b_x, b_y, b_h, b_w, b_a)$ to denote the coordinates of the bounding box for a region. Where b_x, b_y represent the centre coordinates of the bounding box and b_h, b_w, b_a represent the height, width and angle of orientation, respectively. The model trains by minimizing composite loss function $\mathcal{L}_{MTL}$ (defined in Section III-C) by simultaneously optimizing shared and task-specific parameters.

5.1.2 Network Architecture

For an input ECG image I, to extract shared representations across all tasks, we employ ResNet-101 [103] as the backbone network. This architecture comprises 33 sequential residual blocks with identity mapping, where each block contains three convolution layers: a 1×1 convolution, followed by a 3×3 bottleneck layer, and another 1×1 convolution. This backbone architecture transforms the input from the original image space $\mathcal{X}$ into a latent feature space $\mathcal{Z}$. The output representation from the l^{th} residual block is formulated as:

$$\mathbb{Z}_l = F(I_{l-1}, \{\Theta_l\}) + I_{l-1} \tag{5.1}$$

where $F(I_{l-1}, \{\Theta_l\})$ is the residual with parameters Θ_l. Essentially, F computes a residual rather than the output I_l directly as the shortcut connection from the input I_{l-1} is added to the output I_l of the residual block. This bypasses the convolution layers in the module and performs identity mapping with the module output. We remove the last global average pooling and the FC layer of ResNet-101 and take the feature representations, $z \in R^{pxq}$ obtained from the final residual block as the shared features for the tasks.

To address the varying spatial-temporal alignment between 12-lead representations and rhythm strips in standard ECG images, and to accommodate potential tilting in input images, we designed a specialized Region Delineation Network (RDN). This lightweight regression architecture consists of two stacked 3×3 convolutional layers with ReLU activation followed by a prediction output layer. The output layer incorporates a linear layer that predicts both the bounding box coordinates and tilt angle for each region of interest.

The RDN processes shared feature representations z obtained from the backbone network and outputs region-specific bounding box coordinates. Using these coordinates, we implement RROI Align pooling [135] to extract orientation-invariant, region-specific feature maps from the backbone network's output. This technique performs bilinear interpolation to project features from the identified regions of interest into fixed-size feature maps, which are then forwarded to the task-specific classification networks. This approach ensures that regardless of image orientation or alignment variations, the system accurately isolates and processes the distinct ECG regions relevant to each classification task.

For morphological ailment classification, we implement a fully connected network (fc_d) with sigmoid activation ($\sigma(\cdot)$) and corresponding loss function $\mathcal{L}_d$. In parallel, the rhythm classification task utilizes a separate fully connected network (fc_r) with softmax activation and loss function $\mathcal{L}_r$. For a class label l_u, fc_d predicts the confidence score of the presence of label l_u as given in eq. 5.2, and fc_r predicts the class probability across V and is formulated as in eq. 5.3.

$$fc_d(\rho_d(z_n)) = \sigma(w_u^T \rho_d(z_n) + b_u) \tag{5.2}$$

$$fc_r(\rho_r(z_n)) = Softmax((w^r)^T \rho_r(z_n) + b_n) \tag{5.3}$$

where w_u, b_u, w_r and b_r are learnable parameters, σ denotes the sigmoid function. ρ_d and ρ_r represents RRoI align pooling operations. For a test instance x_n, its diagnostic and rhythm prediction is given as $\tilde{y}_n^d$ and $\tilde{y}_n^r$, respectively.

$$\tilde{y}_n^d = \{l_u | fc_d(\rho_d(F_\theta(x)) > 0.5, 1 \leq u \leq U\} \tag{5.4}$$

$$\tilde{y}_n^r = \max_v \{(\exp(w_v^r)^T \rho_r(z_n) / \textstyle\sum_{v=1}^{V} \exp(w_v^r)^T \rho_r(z_n))\} \tag{5.5}$$

During training, the loss obtained from the tasks is combined using uncertainty weights to prevent either of the tasks from dominating the network's performance. The loss functions employed in the proposed framework will be introduced in Section 5.1.3.

5.1.3 Multi-Task Learning Loss

We jointly optimize the three aforementioned tasks through a multi-task learning framework. Each task employs a dedicated loss function: binary cross-entropy loss for the multi-label morphological ailments classification and categorical cross-entropy loss for rhythm classification. For a given input image, $I \in \mathcal{R}^{H \times W \times 3}$ with corresponding ground truth labels y^d,y^r and y^b for disease, rhythm, and region delineation tasks respectively, the learning process integrates these specialized objectives. Specifically, for rhythm

classification, we calculate the categorical cross-entropy loss as formulated in equation 5.6.

$$\mathcal{L}_r(\theta_s, \theta_r) = -\frac{1}{N}\sum_{v=1}^{V}\sum_{n=1}^{N} w_v^r y_{n,c}^r \log P_r \tag{5.6}$$

where N represents the total number of input images, V is the total rhythm classes, and P_r represents the probability scores of the rhythm prediction by the model. θ_s is the shared parameters between the two tasks and θ_r is the task-specific parameters for rhythm classification. w_v^d is the weighting factor that is inversely related to the effective number of samples per class. Following [136], $w_v^r = (1-\alpha)/(1-\alpha^{|v|})$ with $\alpha = (N-1)/N$ and $|v|$ represents total number of samples in class v.

For the multi-label classification of ailments, the problem is converted into a two-class classification problem for each disease label. The loss is obtained by taking the average across each label of each sample. We employed binary cross entropy loss for this, which is given by eq. (5.7). To avoid the effect of class imbalance, we adopted a weighted formulation to achieve a better trade-off with biased sampling.

$$\begin{aligned}\mathcal{L}_d(\theta_s, \theta_d) = -\frac{1}{N}\sum_{u=1}^{U}\sum_{n=1}^{N} w_u^d (y_{n,c}^d \log P_d) \\ +(1-y_{n,c}^d)\log(1-P_d))\end{aligned} \tag{5.7}$$

where P_d is the probability score for ailments classification by the model, U represents the total ailment classes, and θ_d corresponds to task-specific parameters for ailment classification. The loss weight $w_u^d = |Y_{u-}^d|/|Y_{u+}^d|$, and $|Y_{u+}^d|$ and $|Y_{u+}^d|$ denote the positive and negative labels respectively for a class u.

We approach the region delineation task as a regression problem in this work. For optimizing the regression parameters, we implement the $Smooth_{L1}$ loss [137]. This loss function represents a modified version of Mean Squared Error (MSE), offering a smooth transition from quadratic loss behavior for small errors to linear loss behavior for larger errors. The mathematical definition of this regression loss is as follows:

$$\mathcal{L}_b(\theta_s, \theta_b) = \frac{1}{N}\sum_{n=1}^{N} Smooth_{L1}(\hat{y}_n^b, y_n^b) \tag{5.8}$$

and

$$Smooth_{L1}(\hat{y}_n^b, y_n^b) = \begin{cases} 0.5(\hat{y}_n^b - y_n^b)^2/\beta, & \text{if } |\hat{y}_n^b - y_n^b| < \beta. \\ |\hat{y}_n^b - y_n^b| - 0.5\beta, & \text{otherwise} \end{cases} \tag{5.9}$$

where $\hat{y}^b$ is the predicted bounding box coordinates, and the hyper-parameter β acts as a threshold that controls the behaviour of the loss. The loss behaves as MSE if the absolute element-wise error falls below β, and like an $L1$ loss for errors that are larger than β.

In MTL, the network must optimize several loss functions simultaneously, requiring appropriate balancing of each objective's influence during training. While losses could be simply summed across individual tasks, this straightforward approach assumes all task losses have similar magnitudes, an assumption that often doesn't hold. In this work, following our methodology from chapter 4, we implement a dynamic task weighting approach based on homoscedastic uncertainty, as introduced by Kendall et al [128]. Homoscedastic uncertainty represents a type of aleatoric uncertainty that remains constant across all samples within a specific task but varies between different tasks. Assuming the tasks are independently and identically distributed, our multi-task loss function is defined as:

$$\mathcal{L}_{MTL}(\theta_s, \theta_{r,d,b}, \sigma_{r,d,b}) = \frac{1}{2\sigma_d^2}\mathcal{L}_d(\theta_s, \theta_d) + \frac{1}{2\sigma_r^2}\mathcal{L}_r(\theta_s, \theta_r) + \frac{1}{2\sigma_b^2}\mathcal{L}_r(\theta_s, \theta_b) + \log \sigma_d \sigma_r \sigma_b \tag{5.10}$$

where σ_d,$\sigma_r$and σ_b represent uncertainty weights that control the contribution of the respective tasks and are learned by the network during training, it can be observed that the homoscedastic uncertainty terms σ_d, σ_r, σ_b are inversely proportional to model parameters. The term $\log \sigma_d \sigma_r \sigma_b$ acts as a regulizer to avoid the trivial solution.

5.2 Experimental Setup

This section details the experimental methodology used to validate the proposed framework. First, we describe the databases used in this study, including both our internal clinical database and ECG images derived from public ECG datasets. We then present a details of the comparative analysis comprising three main components: ablation studies to assess the contribution of each architectural component, quantitative comparison with existing methods, and qualitative evaluation of the model's predictions. The evaluation metrics used to assess both rhythm and morphological classification performance are then outlined. Finally, we provide implementation details of the proposed framework, including network architecture specifications, training protocols, and hyperparameter settings.

5.2.1 Database

We conducted our experiments using two distinct ECG datasets: the publicly available PTB-XL digital ECG database [91] and an internal private database, AccuRate Cardiocare Clinical ECG (ACCE) database containing clinical ECG images. Since large open-access ECG image datasets are unavailable, we converted digital signals from PTB-XL into ECG images using the process detailed in section 5.2.1.1. We trained our model on these generated PTB-XL images, then evaluated its performance by testing it on the ACCE database. As described in previous chapters, PTB-XL is a comprehensive 12-lead ECG dataset with 10-second recordings capturing various cardiovascular diseases with multiple co-occurring pathologies, plus a substantial portion of healthy ECG data. The pathologies are organized into five diagnostic superclasses: MI, STTC, CD, HYP, and NORM. Table 5.1 presents the statistical breakdown of these diagnostic superclasses. For rhythm classification, the database provides annotations across twelve non-overlapping rhythm categories. Table 5.2 shows the distribution of these rhythm classes, revealing significant imbalance with a long-tail distribution pattern. To prevent the higher-frequency classes from dominating the rhythm classification task, we selected the five most frequent rhythm classes and grouped all remaining classes as 'other.' This approach transforms the task into a six-class classification problem with the following categories: sinus rhythm (SR), atrial fibrillation (AFIB), sinus tachycardia (STACH), sinus arrhythmia (SARRH), sinus bradycardia (SBRAD), and other abnormal rhythms (Othr). We followed the recommended data split for all the experiments in this work.To simulate the variability expected in the input ECG image and to prevent overfitting, particularly for the region delineation task, we applied random rotations within the range of -15 to +15 degrees to all the images. All training and validation were done using the augmented dataset. Additionally, for model evaluation, we applied rotation augmentation to the images within the test fold, and henceforth, we will refer to it as the 'Augmented test fold'.

Table 5.1: PTB-XL dataset overview of diagnostic superclasses

Diagnostic Super class	Description	No. of records	Percentage (%)
NORM	Normal ECG	9528	34.24
MI	Myocardial Infarction	5486	19.72
HYP	Hypertrophy	2655	9.54
CD	Conduction Disturbance	4907	17.63
STTC	ST/T change	5250	18.87

AccuRate Cardiocare Clinical ECG (ACCE) The database comprises a collection of deidentified ECG images, totalling 10977, collected from patients who visited Accurate Cardiocare in the state of Assam, India, between November 2022 and July 2023. These ECGs were recorded as standard 12-lead recordings

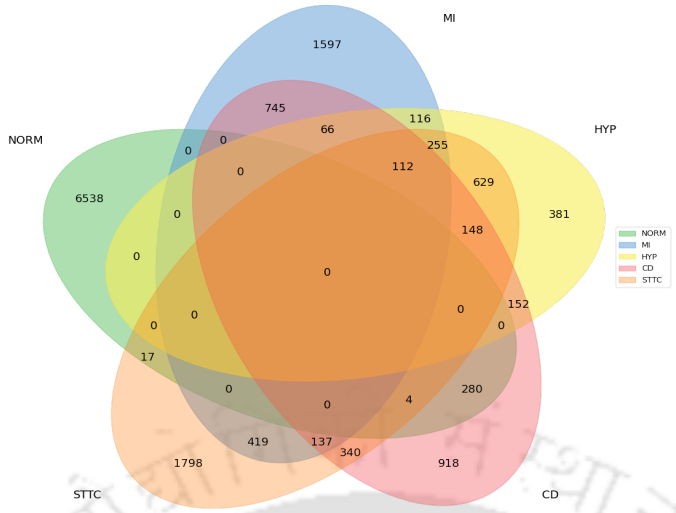

Figure 5.3: Venn Diagram depicting multi-label distribution of diagnostic classes in training set of PTB-XL.

Table 5.2: PTB-XL dataset overview of rhythm classes

Diagnostic Super class	Description	No. of records	Percentage (%)
SR	sinus rhythm	16782	79.12
AFIB	atrial fibrillation	1514	7.14
STACH	sinus tachycardia	826	3.89
SARRH	sinus arrhythmia	772	3.64
SBRAD	sinus bradycardia	637	3.00
PACE	normal functioning artificial pacemaker	296	1.40
SVARR	supraventricular arrhythmia	157	0.74
BIGU	bigeminal pattern	82	0.39
AFLT	atrial flutter	73	0.34
SVTAC	supraventricular tachycardia	27	0.13
PSVT	paroxysmal supraventricular tachycardia	24	0.11
TRIGU	trigeminal pattern	20	0.09

sampled at 500 Hz for 10 s. For these images, the ground-truth annotations were performed by a panel of cardiologists associated with Accurate Cardiocare. This database serves exclusively as a testing dataset, with the specific purpose of evaluating and assessing the generalization performance of the proposed method.

5.2.1.1 Preprocessing

For ECG records from PTB XL, the digital signals are first pre-processed to remove baseline wander using a moving average filter [93]. Non-local means (NLM) filter [94] is used to remove muscle artefacts and high-frequency noises. Further, the signals are normalized using Z-score normalization. The filtered signals are then undergone quality assessment based on [95]. The standard and most commonly used format of printed ECG contains four columns and three rows representing each of the 12 leads in a standard 12-lead

ECG system. Considering a standard-sized paper with a recording speed of 25 mm/s, each column will have three leads of 2.5-second intervals of a continuous 10-second record [9]. A rhythm strip of either lead I, II or III of 10-second continuous recordings is present for rhythm analysis. We employed a python library *ECG plot* and modified it to generate ECG images in standard format from the digital signals. Each record from the database was converted into four ECG images as the records are of 10s length. Each of the four images represents a consecutive 2.5 sec of continuous 10-sec long signal with a common 10-sec long rhythm strip below the 12-lead arrangement.

5.2.2 Comparative Analysis

We assess the effectiveness of our proposed MTL framework in comparison to single-task counterparts for the primary tasks. We trained separate models for each task, which acted as benchmark models for comparison. Furthermore, an investigation is conducted to compare the uncertainty weighting approach against the equal weight loss strategy in the context of MTL.

- **Ablation Study:**

  To evaluate our model's design, we conducted an ablation study examining each component's contribution. We established a *baseline* consisting of our proposed architecture without the RDN and RROI align layer. We then systematically assessed performance improvements by first incorporating RDN into this baseline, followed by analyzing how different pooling operations affect our framework's overall effectiveness.

- **Qualitative Analysis:**

  To better understand the network's predictions and the characteristics of the input regions triggering a particular prediction, we employed Saliency maps [111]. It is a method to obtain post-hoc interpretation of network output using input feature attribution. It highlights the parts of the input region that have a major influence on the network's predictions. By identifying these influential regions, we can verify whether the network focuses on clinically relevant ECG segments, providing validation that the model's focus aligns with established medical knowledge. Additionally, these visualizations enhance the interpretability of our deep learning approach, making the automated diagnostic process more transparent for clinical practitioners.

- **Quantitative Comparison:**

  We conducted a comparative analysis of our proposed approach against four leading state-of-the-art

methods for ECG image analysis: Gliner et al. [76], Sangha et al. [77], Abubaker et al. [75], and Liu et al. [138]. Among these methods, Gliner and Sangha's approaches provide multi-label classification solutions, allowing the assignment of multiple diagnostic labels to a single ECG. To ensure fair and meaningful comparisons across all tasks, we standardized the evaluation framework by adjusting the loss criterion for each competing method to match our proposed approach. This methodological alignment enables direct performance comparisons while maintaining the integrity of each method's underlying architecture and approach.

5.2.3 Evaluation Metrics

To evaluate the performance of the models, we employ five common quantitative metrics, including accuracy, F1 score, specificity, sensitivity and area under receiver operating characteristics curve (AUROC) [97] and precision-recall curve (PRAUC) [98]. The PRAUC metric, which ranges from 0 to 1, provides a quantitative measure of classifier performance. A PR-AUC value of 1 signifies a perfect classifier that achieves the highest possible precision and recall. We take the weighted average of these metrics to represent the overall performance. For multi-label classification, we additionally employ two example-based metrics, namely, hamming loss and coverage rate, which are defined in eq.(5.11) and eq.(5.12), respectively. Hamming Loss evaluates the number of misclassifications in instance-label pairs, the smaller its value, the better the model performance.

$$Hloss = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{U}|\tilde{Y}_n \Delta Y_n| \tag{5.11}$$

where Δ represents symmetric difference between two sets.

Coverage evaluates the average depth to enumerate over the ranked list of predicted label scores to cover all the positive ground truth labels of an instance. This relates to precision at perfect sensitivity. The smaller the coverage, the better the model performance.

$$\textit{coverage} = \frac{1}{N}\sum_{n=1}^{N}\max_{j:\tilde{s}_{nj}=1} rank_{nj} \tag{5.12}$$

where $rank_{ij} = |k : \tilde{s}_{nk} \geq \tilde{s}_{nj}|$ and $\tilde{s} \in R^{NXU}$ represents score associated with each label.

For evaluating the performance of RDN, we utilize four commonly used segmentation metrics: Dice similarity coefficient (DSC), positive predictive value (PPV), Intersection over union (IOU) and Hausdorff distance (HD). The positive predictive value is defined as the ratio of true positives out of all positive outcomes. The Hausdorff distance quantifies the maximum local separation between the two surfaces. The IoU quantifies the overlap by comparing the intersection area of predicted and ground truth regions to their combined

area.

5.2.4 Implementation Details

We developed our implementation in Python utilizing PyTorch (1.5.1) on a high-performance workstation configured with an Intel Xeon 32-core CPU, 64GB RAM, and NVIDIA Tesla P100 GPUs. The composite loss function was optimized via the Adam optimizer with AMSGrad variant, initialized at a learning rate of 0.001 and processing mini-batches of 16 images concurrently. The training protocol incorporated a dynamic learning rate scheduler that applied a 0.2× reduction factor when validation metrics plateaued for ten consecutive epochs. To mitigate overfitting risk, we implemented an early-stopping criterion that terminated the optimization procedure after 25 successive epochs without validation loss improvement. The model checkpoint exhibiting minimum validation loss during the optimization trajectory was preserved for subsequent experimental evaluation and analysis.

5.3 Results and Discussions

In this section, we present performance results of our framework for cardiac rhythm and ailments classification from ECG images. The training process utilized preprocessed PTB-XL dataset signals converted to images using the methods described in section 5.2.1.1. The images thus obtained are split into the train (fold 1-8), validation (fold 9), and test (fold 10) sets, as per database split recommendations in [91]. The training set, including the augmented variant, consisted of 119120 ECG images, while the validation set consisted of 14888 images. Fig 5.3 gives patient-specific distribution of multi-label diagnostic classes in the train fold. For morphological abnormality classification (Task 1), the test set from PTB XL (fold 10) comprised of 1724 images from MI, 1572 from CD, 928 from HYP, 1936 from STTC and 3432 for NORM class. Furthermore, the augmented test set incorporated oriented variant of each image from the test set.

Table 5.3: Classification performance of the proposed network on both of the tasks.

	Multi-Label Morphological Abnormalities					Multi-class Rhythmic Abnormalities				
Database	HLoss ↓	Coverage ↓	F1 Score ↑	AUROC ↑	PRAUC ↑	Accuracy↑	Precision ↑	F1 Score ↑	AUROC ↑	PRAUC ↑
PTB-XL (Test-Fold)	0.1293	1.6240	77.28	92.85	84.04	86.83	84.38	85.05	90.20	87.93
ACCE	0.0989	1.1674	75.60	89.75	78.80	89.61	88.71	87.87	95.72	92.39

Table 5.3 presents classification performance for both tasks using metrics described in Section 5.2.3, evaluated on the PTB-XL test set and ACCE database. For metric interpretation, '↓' indicates lower values

are better, while '↑' indicates higher values are preferred. For Task 1, our approach achieved an F1-score of 77.28%, AUROC of 92.85%, and AUPRC of 84.04% on the PTB-XL test set, with Hamming loss of 0.1293 and coverage of 1.6240. Similar results were observed on the ACCE dataset, with F1-score of 75.60%, AUROC of 89.75%, and AUPRC of 78.80%. For Task 2, the network achieved F1 scores of 85.05% and 87.87% on the PTB-XL test set and ACCE dataset, respectively. These encouraging results for both tasks suggest that the proposed method can adequately extract diverse disease-related features and capitalize on task interdependencies, resulting in enhanced generalization and efficiency across the tasks. Fig. 5.4 displays ROC curves for various abnormality classes in both tasks. These curves plot sensitivity (true positive rate) against 1-specificity (false positive rate) across different decision thresholds, illustrating the balance between correctly identifying positive cases and misclassifying negative ones. The plots show high sensitivity with low false positive rates for both morphological and rhythmic abnormalities, confirming effective model performance. Table 5.4, shows the RDN performance on the PTB-XL test set and augmented test set. Results indicate precise region delineation, with IOU values of 0.984 for the 12-lead alignment region and 0.941 for the rhythm strip.

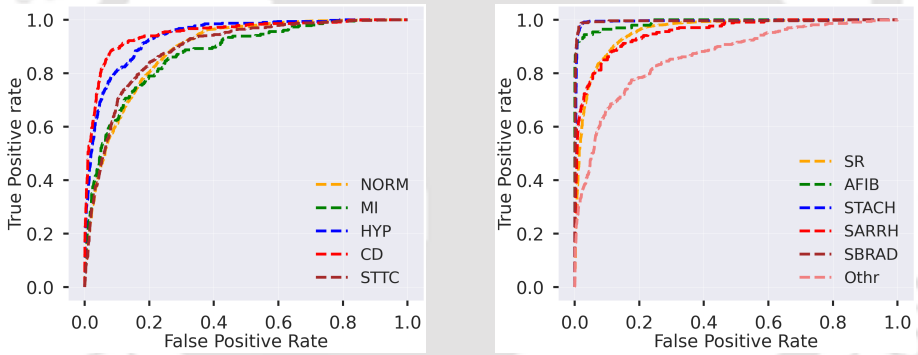

Figure 5.4: The receiver operation characteristic (ROC) curve of (a) Multi-Label Morphological Abnormalities and (b) Multi-class Rhythmic Abnormalities of the proposed network.

Table 5.4: Region detection performance of RDN on test fold and augmented test set of PTX-XL Dataset.

Database	Region	HD (mm) ↓	PPV ↑	IOU ↑	DSC ↑
PTB-XL (Test-Fold)	12-Lead Alignment	2.687	0.986	0.984	0.992
	Rhythm Strip	3.038	0.968	0.941	0.969
PTB-XL (Aug. Test-Fold)	12-Lead Alignment	2.675	0.984	0.982	0.991
	Rhythm Strip	3.119	0.962	0.930	0.963

In Fig. 5.5, we show six ECG images with representative saliency maps. The first two maps correspond

to the morphological abnormalities detection task, one having a single annotation label 'HYP' and the other being a multi-label example having annotations 'MI' and 'STTC'. The third image is an example of the atrial fibrillation ('AFIB') class for the rhythm classification task. As seen from fig 5.5 (a), the importance is scattered across precordial leads V1, V4-V6, the middle portion of the ST-T complex from lead V1, specifically focused around tall R peaks in lead V5. Collectively, these reflect many of the classic criteria for left ventricular hypertrophy (LVH) [139], a subclass of HYP. From Fig 5.5 (b), it can be observed that the model majorly emphasises irregular Q waves and ST segment of lead aVL, lead V2-V3, which is a clinical feature in recognising anteroseptal myocardial infarction [140]. The model also focuses on the ST segment of lead I and lead III, suggesting possible ischemia in infero-lateral region. In Fig 5.5 (c), it can be observed that the region of emphasis is majorly on the P wave region, essentially the absence of P wave; this observation is consistent with the diagnostic criteria of atrial fibrillation [141]. For Fig 5.5 (d), The QRS complex in leads V1 and V2 manifests an 'M'-shaped pattern (rsR') with T-wave inversion observed in lead V1, suggesting a Right Bundle Branch Block (RBBB) [142] , a conduction disorder and the model is effectively highlighting these regions. In case of 5.5 (e), the prominent focus regions are around lead V1, V4, and V5, where leads V1 and V2 demonstrate deep and broad 'QS' or 'rS' complexes characteristic of Left Bundle Branch Block (LBBB) [143], a conduction disorder and tall R peaks are observed in lead V5-V6 indicative of LVH. Fig 5.5 (f) demonstrates an example of sinus bradycardia, showing regular rhythm with consistent P-R intervals and slow heart rate, and the model effectively highlights the diagnostic R-R intervals. As seen from the above examples, these observations from the interpretability analysis of our model's prediction are in parallel with standard clinical ECG interpretation.

Table 5.5 compares single-task and multi-task learning approaches, revealing our network's superior performance over training separate task-specific models. The single-task network for Task 1 achieved an F1 Score of 75.73%, AUROC of 91.46%, and PRAUC of 81.65%, while the single-task network for Task 2 reached 85.04% accuracy and 84.38% weighted average F1-score. Our proposed multi-task network improved Task 1 performance significantly: F1 Score by 1.55%, AUROC by 1.39%, and PRAUC by 2.39%. Similarly, improvements were observed across all metrics for Task 2 compared to its corresponding single-task model. These enhancements highlight the benefits of feature-sharing in multi-task learning. We also compared our approach against a basic loss addition-based multi-task method where task-specific losses receive equal weights. Results in Table 5.5 demonstrate that our uncertainty-weighted multi-task learning approach outperforms the equal-weights method for both tasks. This confirms the effectiveness of uncertainty weights in balancing morphological and rhythm classification tasks, preventing either task from dominating the training process.

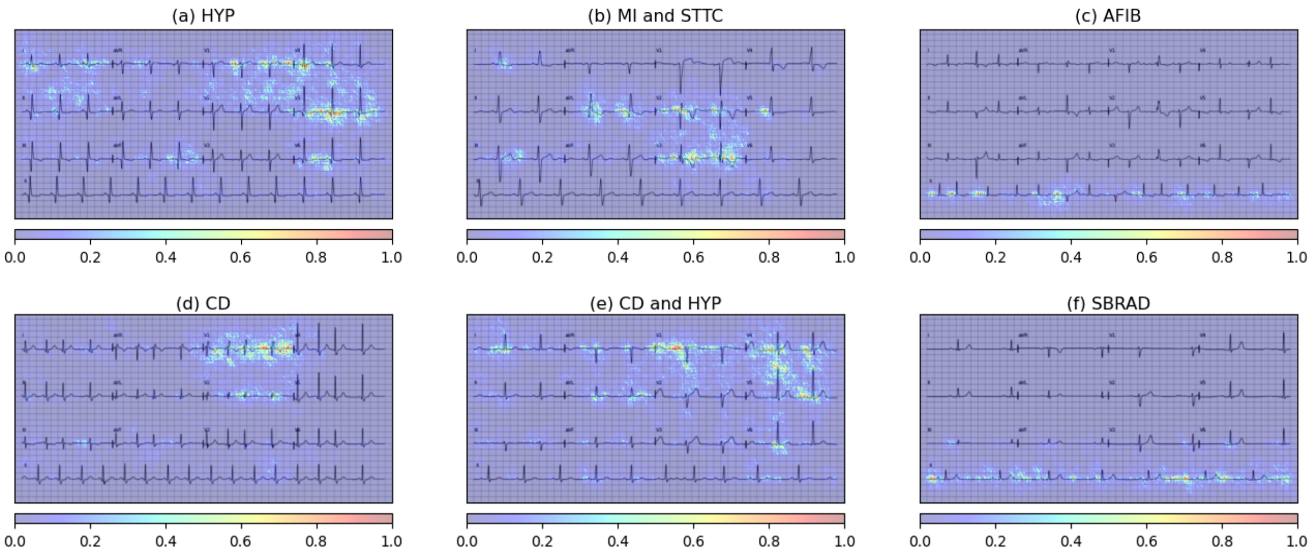

Figure 5.5: Visualization of saliency maps obtained by the proposed method, overlayed over corresponding ECG images.

Table 5.5: Comparative results between single and multi-task learning approaches on PTB XL Test Set

	Multi-Label Morphological Abnormalities					Multi-class Rhythmic Abnormalities				
Method	HLoss ↓	Coverage ↓	F1 Score ↑	AUROC ↑	PRAUC ↑	Accuracy↑	Precision ↑	F1 Score ↑	AUROC ↑	PRAUC ↑
Single Task (Task 1)	0.1300	1.6907	75.73	91.46	81.65	–	–	–	–	–
Single Task (Task 2)	–	–	–	–	–	85.04	83.89	84.38	87.42	86.03
Equal weights MTL	0.1293	1.6767	75.65	91.38	80.98	85.74	84.09	84.81	88.28	86.66
This Work	**0.1293**	**1.6240**	**77.28**	**92.85**	**84.04**	**86.83**	**84.38**	**85.05**	**90.20**	**87.93**

Table 5.6 presents the ablation results. In the baseline configuration, without region information available, we only utilized feature representations from the backbone network for classification. Our complete method outperformed this baseline in both morphological and rhythmic abnormality detection, improving F1 scores by 1.33% and 1.03% respectively on the PTB-XL test set. The addition of RDN facilitates the passing of region-specific information, which greatly enhances task-specific feature representation and leading to improved performance across primary tasks. To evaluate the effectiveness of our feature pooling approach, we conducted an experiment replacing the RROI Align layer with an RROI pool layer. This substitution resulted in decreased classification performance, confirming the superiority of our bilinear interpolation-based RROI Align compared to the max-pooling-based RROI pool method.

We further validate the performance of our approach by comparing it with four state-of-the-art methods on ECG images: Gliner et al. [76], Sangha et al. [77], Abubaker et al. [75], and Liu et al. [138]. Table 5.7 summarises the average classification performance of the comparative analysis on Task 1 and Task

Table 5.6: Ablative study of different components in our proposed method

Method	Multi-Label Morphological Abnormalities					Multi-class Rhythmic Abnormalities				
	HLoss ↓	Coverage ↓	F1 Score ↑	AUROC ↑	PRAUC ↑	Accuracy↑	Precision ↑	F1 Score ↑	AUROC ↑	PRAUC ↑
Baseline	0.1302	1.6600	75.95	91.99	81.88	84.85	83.41	84.02	87.45	85.87
Baseline + RDN + RROI Pool	0.1370	1.6441	76.71	92.83	83.94	86.11	84.17	84.69	88.55	86.80
Baseline + RDN + RROI Align	**0.1293**	**1.6240**	**77.28**	**92.85**	**84.04**	**86.83**	**84.38**	**85.05**	**90.20**	**87.93**

Table 5.7: Comparison Results of our proposed method and other existing state-of-the-art methods for morphological and rhythmic abnormalities classification task on PTB XL Test Set

Database	Method	Multi-Label Morphological Abnormalities					Multi-class Rhythmic Abnormalities				
		HLoss ↓	Coverage ↓	F1 Score ↑	AUROC ↑	PRAUC ↑	Accuracy↑	Precision ↑	F1 Score ↑	AUROC ↑	PRAUC ↑
PTB-XL (Test-Fold)	Abubaker et al. [75]	–	–	–	–	–	82.66	80.05	80.68	85.54	83.19
	Sangha et al. [77]	**0.1134**	1.6338	76.37	92.49	83.65	–	–	–	–	–
	Gliner et al. [76]	0.1277	1.6869	74.10	90.51	79.73	–	–	–	–	–
	Liu et al. [138]	–	–	–	–	–	86.59	83.81	84.21	87.52	85.82
	This Work	0.1293	**1.6240**	**77.28**	**92.85**	**84.04**	**86.83**	**84.38**	**85.05**	**90.20**	**87.93**
ACCE	Abubaker et al. [75]	–	–	–	–	–	83.91	85.09	82.26	91.90	87.69
	Sangha et al. [77]	0.0965	1.1806	70.41	89.72	**78.86**	–	–	–	–	–
	Gliner et al. [76]	**0.0932**	1.1991	71.66	89.21	77.79	–	–	–	–	–
	Liu et al. [138]	–	–	–	–	–	88.21	87.19	85.28	95.63	92.28
	This Work	0.0989	**1.1674**	**75.60**	**89.75**	78.80	**89.61**	**88.71**	**87.87**	**95.72**	**92.39**

2 across PTB XL test set and ACCE dataset. As shown in Table 5.7, our proposed method, which combines multi-task learning with region-specific delineation and homoscedastic uncertainty-based task weighting, achieves superior performance across both datasets. Gliner et al. [76] and Sangha et al. [77] utilized CNN-based architectures for multi-label approaches. Gliner employed a custom CNN with seven convolutional layers followed by a fully connected layer, while Sangha utilized EfficientNet's B3 variant, both showing competitive performance for morphological abnormality detection. These methods performed better in terms of Hamming loss, suggesting fewer overall mistakes across all labels compared to our approach, but they appear more conservative in positive case predictions as indicated by their lower F1 scores. However, our approach demonstrates superior performance in other critical metrics, including F1 score, AUROC, AUPRC, and coverage, indicating better class discrimination, improved performance on imbalanced data, and broader applicability across diverse cases. Abubaker et al. [75], and Liu et al. [138] provided CNN-based multi-class solutions with promising results for Task 2 across both datasets. However, their performance is slightly lower compared to our approach, which incorporates additional context through

region-specific delineation and task-specific homoscedastic uncertainty.

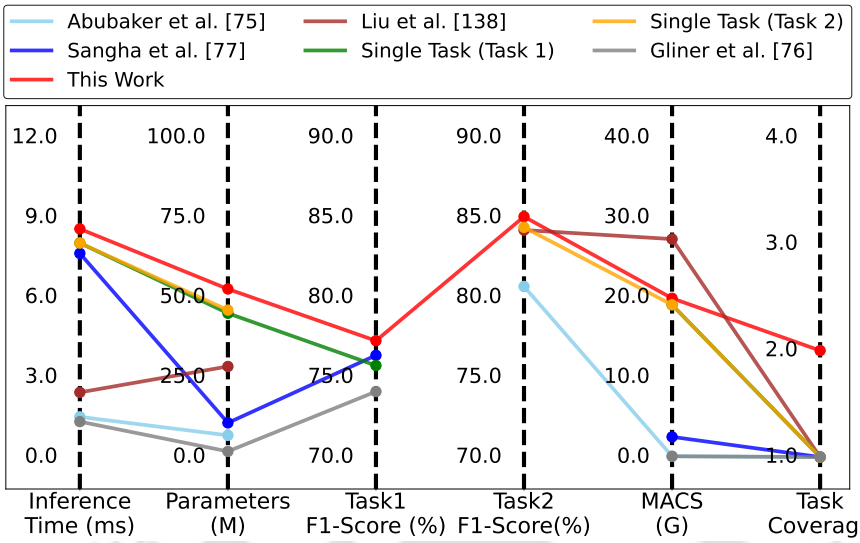

Figure 5.6: Parallel coordinate visualization comparing computational complexity and efficiency trade-offs across methods.

In Fig. 5.6, we present a parallel coordinate plot to analyze the computational complexity of our proposed multi-task network in comparison with the state-of-the-art methods and the single task variants of our proposed approach. The plot illustrates trade-offs between computational demands and model effectiveness, with metrics including number of parameters (in millions, M), inference time (in milliseconds, ms), multiply-accumulate operations (MACs) (in giga, G), task F1-scores, and task coverage. As can be observed from the figure, while the proposed method has higher model parameters (≈53M), it demonstrates superior performance in both tasks. Notably, it simultaneously addresses both objectives, which effectively eliminates the need for deploying multiple separate models. Also, the proposed network achieves an efficient inference speed of 8.5ms, making it suitable for real-time medical deployment.

5.4 Summary

In this work, we developed a multi-task CNN architecture to simultaneously diagnose morphological and rhythmic cardiac ailments from ECG images. The proposed network is trained in an end-to-end manner, which ascertains only application-specific features are learnt. The network utilized homoscedastic uncertainty-based dynamic weighting to balance the tasks and employed task-specific ROIs to distil features for each specific task. We employed a multi-label classification approach for the morphological ailment classification task to accommodate the possibility of cardiac comorbidities. The proposed framework is evaluated on ECG images derived from the PTB-XL dataset and an internal clinical ECG image database.

Experimental results show that our proposed method achieves encouraging performance and outperforms single-task approaches and other state-of-the-art methods. In summary, based on quantitative results, our multi-task learning framework effectively leverages task interdependencies by optimizing task-specific feature representations. The model demonstrates enhanced discriminative power in detecting diverse cardiac abnormalities from ECG images, yielding improved generalization and task-specific performance as evidenced by superior metrics across multiple evaluation criteria. Furthermore, complexity analysis confirms the clinical viability of our approach, with the network achieving an efficient inference speed, making it well-suited for real-time medical deployment in practical healthcare settings.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

6

Conclusions

Contents

CVDs pose a critical global health challenge, causing nearly one-third of all deaths worldwide. This significant healthcare burden, coupled with the growing volume and complexity of ECG data, necessitates reliable automated diagnostic support systems to assist clinicians in timely and accurate diagnosis. While deep learning approaches have shown promising results in automated ECG analysis, their translation to clinical practice faces several fundamental challenges. Current methods typically adopt a purely data-driven approach, which might overlook crucial clinical knowledge and diagnostic patterns. Furthermore, these approaches lack essential characteristics of clinical diagnosis. First, most methods provide only deterministic predictions without quantifying their uncertainty - a crucial aspect in clinical settings where confidence in predictions significantly influences decision-making, particularly in complex cases where misdiagnosis could have severe consequences. Second, cardiac conditions often present with comorbidities, with studies indicating that over 60% of cardiovascular patients have multiple concurrent conditions. This necessitates systems capable of detecting multiple abnormalities while respecting their inherent relationships, rather than treating them as independent flat classifications. Third, many cardiac conditions follow natural diagnostic hierarchies, where broad categories encompass specific subtypes, requiring frameworks that maintain these hierarchical relationships for consistent diagnosis. This thesis addresses these fundamental limitations by developing frameworks that systematically integrate domain expertise into deep learning models. Our approaches ensure reliable uncertainty quantification, handle multi-label scenarios while preserving condition relationships, and respect diagnostic hierarchies, advancing towards more clinically applicable automated ECG interpretation systems that can assist in addressing the growing global burden of cardiovascular disease. The summary of each chapter of the thesis is given as follows.

Chapter 1 provides the clinical and technical foundation for this thesis, beginning with an overview of cardiovascular physiology and electrocardiography principles. It discusses the genesis and clinical significance of ECG signals, exploring how various cardiac conditions manifest in ECG recordings through characteristic pathological patterns. Through a systematic review of automated ECG diagnostic systems, we identify several crucial gaps in current approachess. The chapter concludes by outlining these research gaps and establishing the motivation for the current research. In the following paragraphs, we summarize our contributions that address these fundamental limitations, followed by a comprehensive summary of our findings.

Chapter 2 presented the CTAT framework, a novel approach to improve MI diagnosis from ECG signals by incorporating clinical context knowledge of diagnostically salient morphological regions into a deep learning framework. The framework addresses a fundamental limitation of current deep learning approaches by guiding models to focus on diagnostically significant ECG regions through knowledge

distillation between ECG delineation and MI detection tasks. The architecture consists of three key components: a teacher network based on U-Net with attention and residual connections that captures morphological features through ECG delineation, a student network based on ResNet with spatial attention optimized for MI classification, and a CTAT module that facilitates the transfer of learned attention patterns between these networks. This knowledge transfer enables the student network to focus on clinically relevant ECG segments, mimicking expert analysis of specific waveform characteristics during MI diagnosis. Comprehensive evaluation on multiple 12-lead ECG databases demonstrates the framework's superior performance over existing methods, while attention visualization confirms its focus on diagnostically relevant ECG regions.

Chapter 3 extended this concept through MTPCHi-MI framework. It is a unified multi-task framework for context-aware hierarchical classification of MI. It employs two auxiliary tasks corresponding to ECG delineation and ECG reconstruction to enable a more complex shared latent representation of ECG data, capturing additional morphological context along with physics-informed signal dynamics.The framework features an encoder-decoder architecture where a shared encoder learns common representations across three complementary tasks: hierarchical MI classification, ECG delineation, and physics-informed ECG reconstruction. Twin parallel decoders based on U-Net with attention and residual connections handle the auxiliary tasks, while an LSTM-based decoder performs hierarchical MI classification through its autoregressive nature, enabling sequential predictions that respect the natural progression from MI detection to localization. The framework introduces a novel hierarchical loss function that enforces the inherent relationships between different levels of MI diagnosis, while the physics-informed reconstruction task ensures the learned representations preserve the underlying characteristics of 12-lead ECG signals. This comprehensive approach enables more robust and clinically consistent MI diagnosis by jointly leveraging morphological, physical, and hierarchical aspects of ECG interpretation.

Chapter 4 addressed the critical need for reliable uncertainty estimation in clinical deployment through our uncertainty-aware multi-label cardiac diagnostic framework. The framework uniquely combines variational inference with correlation-aware multivariate sigmoid to provide uncertainty estimates for multi-label cardiac abnormality detection. At its core, the architecture employs a variational autoencoder with Gaussian mixture latent representation to learn the underlying ECG signal distribution, enabling uncertainty estimation through analysis of the posterior distribution in latent space. The framework incorporates a correlation-aware multivariate sigmoid function with a learnable interaction matrix that captures interdependencies between cardiac conditions. The framework not only provides accurate predictions but also reliably identifies cases requiring additional clinical review, particularly when encountering ECG patterns

that deviate from the training distribution through effective out-of-distribution detection capabilities. Comprehensive evaluation demonstrates the framework's effectiveness in maintaining high diagnostic accuracy while providing reliable uncertainty estimates crucial for clinical deployment.

Finally, Chapter 5 tackled the challenge of paper ECG analysis through a region-aware multi-task learning framework. While previous chapters focused on enhancing digital ECG analysis through domain knowledge integration and uncertainty quantification, this chapter bridges the gap between digital algorithms and the paper ECG format that remains prevalent in many healthcare settings. Drawing inspiration from clinical practice, where experts analyze different regions of ECG prints for specific diagnostic purposes, we developed a region-aware multi-task learning framework. The architecture effectively translates this structured interpretation process into an automated system through three key components: a ResNet backbone for feature extraction, region-based pooling for separate analysis of rhythm strips and 12-lead sections, and task-specific branches for comprehensive cardiac diagnosis. The framework's effectiveness, demonstrated through extensive evaluation on both public datasets and internal clinical records, establishes a practical solution for automated paper ECG interpretation. This contribution complements our previous frameworks by ensuring that automated ECG analysis can be deployed across diverse clinical settings, regardless of whether they use digital or paper ECG formats.

Together, these contributions establish new paradigms for developing clinically reliable automated ECG interpretation systems, paving the way for their practical deployment in healthcare settings.

6.1 Future Directions

While this thesis presents significant advances in automated ECG interpretation, several promising directions warrant further investigation:

(i) **Physics-Informed Cardiac Modeling**: The integration of cardiac electrophysiological principles with deep learning presents an exciting direction for future research. Development of physics-informed neural networks that explicitly incorporate electrophysiological models could provide more biologically plausible representations and better interpretability. This approach could enable joint learning of ECG characteristics and underlying physiological parameters, potentially improving the accuracy of cardiac abnormality localization. Such frameworks could also help in solving the inverse problem of determining cardiac electrical activity from surface ECG measurements, providing deeper insights into cardiac function and pathology.

(ii) **Edge AI and Real-time Systems**: The translation of automated ECG interpretation to real-world

clinical settings necessitates efficient, real-time processing capabilities. Future research should focus on developing lightweight architectures suitable for deployment on resource-constrained devices while maintaining diagnostic accuracy. This includes investigating model compression techniques, optimizing inference speed, and developing edge-cloud collaborative frameworks for efficient processing. Such systems could enable continuous cardiac monitoring with real-time warning capabilities, particularly valuable for early detection of acute cardiac events in both hospital and remote monitoring settings.

(iii) **Multi-modal Integration**: Future research should explore the integration of multiple cardiac diagnostic modalities to provide more comprehensive and robust diagnosis. This includes developing frameworks that can effectively combine ECG data with other cardiac imaging modalities (e.g., echocardiography, cardiac MRI), clinical metadata, and patient history. Advanced attention mechanisms for cross-modal feature learning and strategies for handling missing modalities, a common scenario in clinical settings, will be crucial. Such multi-modal approaches could provide more context-aware and personalized diagnostic capabilities.

(iv) **Multi-label Learning with label corelation**: The complex nature of cardiac comorbidities necessitates more advanced approaches to multi-label classification. Future work should focus on developing methods that can effectively model disease co-occurrence patterns and capture complex label dependencies, potentially through graph-based approaches. This includes developing hierarchical multi-label classification frameworks with uncertainty quantification and mechanisms for dynamically adapting label relationships based on patient-specific factors.

(v) **Interpretable AI and Causal Learning**:

While current approaches provide some level of interpretability, future research should focus on developing more sophisticated explanation methods that align with clinical reasoning. This includes investigating causal learning approaches to understand relationships between ECG features and cardiac conditions, moving beyond mere correlation to capture underlying physiological mechanisms. Such advances could help in generating clinically actionable insights and improving trust in automated diagnostic systems.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

References

[1] P. Bizopoulos and D. Koutsouris, "Deep learning in cardiology," *IEEE reviews in biomedical engineering*, vol. 12, pp. 168–193, 2018.

[2] P. Elias, S. S. Jain, T. Poterucha, M. Randazzo, F. Lopez Jimenez, R. Khera, M. Perez, D. Ouyang, J. Pirruccello, M. Salerno *et al.*, "Artificial intelligence for cardiovascular care—part 1: advances: Jacc review topic of the week," *Journal of the American College of Cardiology*, vol. 83, no. 24, pp. 2472–2486, 2024.

[3] J. L. Anderson and D. A. Morrow, "Acute myocardial infarction," *New England Journal of Medicine*, vol. 376, no. 21, pp. 2053–2064, 2017.

[4] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, pp. 65–69, 2019.

[5] "Cardiovascular diseases, WHO," https://www.who.int/health-topics/hypertension/cardiovascular-diseases, [Accessed 24-12-2023].

[6] H. Ueshima, A. Sekikawa, K. Miura, T. C. Turin, N. Takashima, Y. Kita, M. Watanabe, A. Kadota, N. Okuda, T. Kadowaki *et al.*, "Cardiovascular disease and risk factors in asia: a selected review," *Circulation*, vol. 118, no. 25, pp. 2702–2709, 2008.

[7] F. Lopez-Jimenez, Z. Attia, A. M. Arruda-Olson, R. Carter, P. Chareonthaitawee, H. Jouni, S. Kapa, A. Lerman, C. Luong, J. R. Medina-Inojosa *et al.*, "Artificial intelligence in cardiology: present and future," in *Mayo Clinic Proceedings*, vol. 95, no. 5. Elsevier, 2020, pp. 1015–1039.

[8] J. J. Thiagarajan, D. Rajan, S. Katoch, and A. Spanias, "Ddxnet: a deep learning model for automatic interpretation of electronic health records, electrocardiograms and electroencephalograms," *Scientific reports*, vol. 10, no. 1, p. 16428, 2020.

[9] P. Kligfield, L. S. Gettes, J. J. Bailey, R. Childers, B. J. Deal, E. W. Hancock, G. Van Herpen, J. A. Kors, P. Macfarlane, D. M. Mirvis *et al.*, "Recommendations for the standardization and interpretation of the electrocardiogram: part i: the electrocardiogram and its technology a scientific statement from the american heart association electrocardiography and arrhythmias committee, council on clinical cardiology; the american college of cardiology foundation; and the heart rhythm society endorsed by the international society for computerized electrocardiology," *Journal of the American College of Cardiology*, vol. 49, no. 10, pp. 1109–1127, 2007.

[10] S. M. Salerno, P. C. Alguire, and H. S. Waxman, "Competency in interpretation of 12-lead electrocardiograms: a summary and appraisal of published evidence," *Annals of Internal Medicine*, vol. 138, no. 9, pp. 751–760, 2003.

[11] M. Bickerton and A. Pooler, "Misplaced ecg electrodes and the need for continuing training," *British Journal of Cardiac Nursing*, vol. 14, no. 3, pp. 123–132, 2019.

[12] L. Zou, Z. Huang, X. Yu, J. Zheng, A. Liu, and M. Lei, "Automatic detection of congestive heart failure based on multiscale residual unet++: From centralized learning to federated learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–13, 2022.

[13] C. Han, Y. Zhou, W. Que, Z. Li, and L. Shi, "An overview of algorithms for myocardial infarction diagnostics using ecg signals: Advances and challenges," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[14] S. Ansari, N. Farzaneh, M. Duda, K. Horan, H. B. Andersson, Z. D. Goldberger, B. K. Nallamothu, and K. Najarian, "A review of automated methods for detection of myocardial ischemia and infarction using electrocardiogram and electronic health records," *IEEE reviews in biomedical engineering*, vol. 10, pp. 264–298, 2017.

[15] P. S. Choudhary and S. Dandapat, "Morphology-aware ecg diagnostic framework with cross-task attention transfer for improved myocardial infarction diagnosis," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[16] M. Barandas, L. Famiglini, A. Campagner, D. Folgado, R. Simão, F. Cabitza, and H. Gamboa, "Evaluation of uncertainty quantification methods in multi-label classification: A case study with automatic diagnosis of electrocardiogram," *Information Fusion*, vol. 101, p. 101978, 2024.

[17] A. C. Guyton, *Text book of medical physiology*. China, 2006.

[18] J. E. Hall and M. E. Hall, *Guyton and Hall Textbook of Medical Physiology E-Book: Guyton and Hall Textbook of Medical Physiology E-Book*. Elsevier Health Sciences, 2020.

[19] G. D. Clifford, F. Azuaje, P. McSharry *et al.*, *Advanced methods and tools for ECG data analysis*. Artech house Boston, 2006, vol. 10.

[20] A. L. Goldberger, Z. D. Goldberger, and A. Shvilkin, *Clinical Electrocardiography: A Simplified Approach: Clinical Electrocardiography: A Simplified Approach E-Book*. Elsevier Health Sciences, 2017.

[21] K. Thygesen, J. S. Alpert, A. S. Jaffe, B. R. Chaitman, J. J. Bax, D. A. Morrow, H. D. White, and E. G. on behalf of the Joint European Society of Cardiology (ESC)/American College of Cardiology (ACC)/American Heart Association (AHA)/World Heart Federation (WHF) Task Force for the Universal Definition of Myocardial Infarction, "Fourth universal definition of myocardial infarction (2018)," *Circulation*, vol. 138, no. 20, pp. e618–e651, 2018.

[22] W. J. Brady and F. Morris, "Electrocardiographic abnormalities encountered in acute myocardial infarction," *Emergency Medicine Journal*, vol. 17, no. 1, pp. 40–45, 2000.

[23] J. V. Nable and W. Brady, "The evolution of electrocardiographic changes in st-segment elevation myocardial infarction," *The American journal of emergency medicine*, vol. 27, no. 6, pp. 734–746, 2009.

[24] A. L. Goldberger and E. Gold-berger, "Clinical electrocardiography, a simplified approach," *Critical Care Medicine*, vol. 9, no. 12, pp. 891–892, 1981.

[25] K. Thygesen, J. S. Alpert, A. S. Jaffe, M. L. Simoons, B. R. Chaitman, and H. D. White, "Third universal definition of myocardial infarction," *circulation*, vol. 126, no. 16, pp. 2020–2035, 2012.

[26] I. Shimizu and T. Minamino, "Physiological and pathological cardiac hypertrophy," *Journal of molecular and cellular cardiology*, vol. 97, pp. 245–262, 2016.

[27] L. M. Bacharova, D. Schocken, E. H Estes, and D. Strauss, "The role of ecg in the diagnosis of left ventricular hypertrophy," *Current cardiology reviews*, vol. 10, no. 3, pp. 257–261, 2014.

[28] C. J. McLeod, M. J. Ackerman, R. A. Nishimura, A. J. Tajik, B. J. Gersh, and S. R. Ommen, "Outcome of patients with hypertrophic cardiomyopathy and a normal electrocardiogram," *Journal of the american College of Cardiology*, vol. 54, no. 3, pp. 229–233, 2009.

[29] T. Kerola, A. Eranti, A. L. Aro, M. A. Haukilahti, A. Holkeri, M. J. Junttila, T. V. Kenttä, H. Rissanen, E. Vittinghoff, P. Knekt *et al.*, "Risk factors associated with atrioventricular block," *JAMA Network Open*, vol. 2, no. 5, pp. e194 176–e194 176, 2019.

[30] P. Francia, C. Balla, F. Paneni, and M. Volpe, "Left bundle-branch block—pathophysiology, prognosis, and clinical management," *Clinical Cardiology: An International Indexed and Peer-Reviewed Journal for Advances in the Treatment of Cardiovascular Disease*, vol. 30, no. 3, pp. 110–115, 2007.

[31] T. Ikeda, "Right bundle branch block: current considerations," *Current Cardiology Reviews*, vol. 17, no. 1, pp. 24–30, 2021.

[32] Z. Ebrahimi, M. Loni, M. Daneshtalab, and A. Gharehbaghi, "A review on deep learning methods for ecg arrhythmia classification," *Expert Systems with Applications: X*, vol. 7, p. 100033, 2020.

[33] M. V. Perez, F. E. Dewey, R. Marcus, E. A. Ashley, A. A. Al-Ahmad, P. J. Wang, and V. F. Froelicher, "Electrocardiographic predictors of atrial fibrillation," *American heart journal*, vol. 158, no. 4, pp. 622–628, 2009.

[34] P. Macfarlane, B. Devine, S. Latif, S. McLaughlin, D. Shoat, and M. Watts, "Methodology of ecg interpretation in the glasgow program," *Methods of information in medicine*, vol. 29, no. 04, pp. 354–361, 1990.

[35] C. Zywietz, D. Borovsky, G. Gotsch, and G. Joseph, "Methodology of ecg interpretation in the hannover program," *Methods of information in medicine*, vol. 29, no. 04, pp. 375–385, 1990.

[36] J. Van Bemmel, J. Kors, and G. Van Herpen, "Methodology of the modular ecg analysis system means," *Methods of information in medicine*, vol. 29, no. 04, pp. 346–353, 1990.

[37] G. Sahu and K. C. Ray, "An efficient method for detection and localization of myocardial infarction," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2021.

[38] L. Sharma, R. Tripathy, and S. Dandapat, "Multiscale energy and eigenspace approach to detection and localization of myocardial infarction," *IEEE transactions on biomedical engineering*, vol. 62, no. 7, pp. 1827–1837, 2015.

[39] S. Padhy and S. Dandapat, "Third-order tensor based analysis of multilead ecg for classification of myocardial infarction," *Biomedical Signal Processing and Control*, vol. 31, pp. 71–78, 2017.

[40] S. S. Al-Zaiti, C. Martin-Gill, J. K. Zègre-Hemsey, Z. Bouzid, Z. Faramand, M. O. Alrawashdeh, R. E. Gregg, S. Helman, N. T. Riek, K. Kraevsky-Phillips *et al.*, "Machine learning for ecg diagnosis and risk stratification of occlusion myocardial infarction," *Nature Medicine*, vol. 29, no. 7, pp. 1804–1813, 2023.

[41] O. Yildirim, U. B. Baloglu, R.-S. Tan, E. J. Ciaccio, and U. R. Acharya, "A new approach for arrhythmia classification using deep coded features and lstm networks," *Computer methods and programs in biomedicine*, vol. 176, pp. 121–133, 2019.

[42] E. Prabhakararao and S. Dandapat, "Myocardial infarction severity stages classification from ecg signals using attentional recurrent neural network," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8711–8720, 2020.

[43] J. Yu, J. Gao, N. Wang, P. Feng, B. Zhou, and Z. Wang, "Spa-tem mi: A spatial-temporal network for detecting and locating myocardial infarction," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[44] D. Jyotishi and S. Dandapat, "An attentive spatio-temporal learning-based network for cardiovascular disease diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023.

[45] U. Maji, M. Mitra, and S. Pal, "Imposed target based modification of taguchi method for feature optimisation with application in arrhythmia beat detection," *Expert Systems with Applications*, vol. 56, pp. 268–281, 2016.

[46] W. Zeng and C. Yuan, "Ecg arrhythmia classification based on variational mode decomposition, shannon energy envelope and deterministic learning," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 10, pp. 2963–2988, 2021.

[47] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE transactions on biomedical engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.

[48] M. Arif, I. A. Malagore, and F. A. Afsar, "Detection and localization of myocardial infarction using k-nearest neighbor classifier," *Journal of medical systems*, vol. 36, pp. 279–289, 2012.

[49] D. Romero, M. Ringborn, P. Laguna, and E. Pueyo, "Detection and quantification of acute myocardial ischemia by morphologic evaluation of qrs changes by an angle-based method," *Journal of electrocardiology*, vol. 46, no. 3, pp. 204–214, 2013.

[50] M. Hotradat, K. Balasundaram, S. Masse, K. Nair, K. Nanthakumar, and K. Umapathy, "Empirical mode decomposition based ecg features in classifying and tracking ventricular arrhythmias," *Computers in Biology and Medicine*, vol. 112, p. 103379, 2019.

[51] M. T. Nguyen, B. Van Nguyen, and K. Kim, "Shockable rhythm diagnosis for automated external defibrillators using a modified variational mode decomposition technique," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3037–3046, 2017.

[52] U. Maji, M. Mitra, and S. Pal, "Characterization of cardiac arrhythmias by variational mode decomposition technique," *Biocybernetics and Biomedical Engineering*, vol. 37, no. 3, pp. 578–589, 2017.

[53] S. M. Qaisar, S. I. Khan, K. Srinivasan, and M. Krichen, "Arrhythmia classification using multirate processing metaheuristic optimization and variational mode decomposition," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 1, pp. 26–37, 2023.

[54] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[55] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE transactions on biomedical engineering*, vol. 63, no. 3, pp. 664–675, 2015.

[56] D. U. Jeong and K. M. Lim, "Convolutional neural network for classification of eight types of arrhythmia using 2d time–frequency feature map from standard 12-lead electrocardiogram," *Scientific reports*, vol. 11, no. 1, p. 20396, 2021.

[57] B. Król-Józaga, "Atrial fibrillation detection using convolutional neural networks on 2-dimensional representation of ecg signal," *Biomedical Signal Processing and Control*, vol. 74, p. 103470, 2022.

[58] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ecg features for screening paroxysmal atrial fibrillation patients," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2095–2104, 2017.

[59] Z. Lipton, "Learning to diagnose with lstm recurrent neural networks," *arXiv preprint arXiv:1511.03677*, 2015.

[60] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997.

[61] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[62] J. Wang, "A modified state refinement gru for atrial fibrillation and premature ventricular contraction signals detection," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2024.

[63] M. Dey, N. Omar, and M. A. Ullah, "Temporal feature-based classification into myocardial infarction and other cvds merging cnn and bi-lstm from ecg signal," *IEEE Sensors Journal*, vol. 21, no. 19, pp. 21 688–21 695, 2021.

[64] C. Chen, Z. Hua, R. Zhang, G. Liu, and W. Wen, "Automated arrhythmia classification based on a combination network of cnn and lstm," *Biomedical Signal Processing and Control*, vol. 57, p. 101819, 2020.

[65] G. Petmezas, K. Haris, L. Stefanopoulos, V. Kilintzis, A. Tzavelis, J. A. Rogers, A. K. Katsaggelos, and N. Maglaveras, "Automated atrial fibrillation detection using a hybrid cnn-lstm network on imbalanced ecg datasets," *Biomedical Signal Processing and Control*, vol. 63, p. 102194, 2021.

[66] W. Liu, F. Wang, Q. Huang, S. Chang, H. Wang, and J. He, "Mfb-cbrnn: A hybrid network for mi detection using 12-lead ecgs," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 503–514, 2019.

[67] C. Han and L. Shi, "Ml–resnet: A novel network to detect and locate myocardial infarction using 12 leads ecg," *Computer methods and programs in biomedicine*, vol. 185, p. 105138, 2020.

[68] C. Han, J. Sun, Y. Bian, W. Que, and L. Shi, "Automated detection and localization of myocardial infarction with interpretability analysis based on deep learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023.

[69] T. Dissanayake, T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Dconv-lstm-net: A novel architecture for single and 12-lead ecg anomaly detection." *IEEE Sensors Journal*, 2023.

[70] S. Yang, C. Lian, Z. Zeng, B. Xu, J. Zang, and Z. Zhang, "A multi-view multi-scale neural network for multi-label ecg classification," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023.

[71] Q.-J. Lv, H.-Y. Chen, W.-B. Zhong, Y.-Y. Wang, J.-Y. Song, S.-D. Guo, L.-X. Qi, and C. Y.-C. Chen, "A multi-task group bi-lstm networks application on electrocardiogram classification," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1–11, 2019.

[72] E. Prabhakararao and S. Dandapat, "Atrial fibrillation burden estimation using multi-task deep convolutional neural network," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 12, pp. 5992–6002, 2022.

[73] P. Hao, X. Gao, Z. Li, J. Zhang, F. Wu, and C. Bai, "Multi-branch fusion network for myocardial infarction screening from 12-lead ecg images," *Computer methods and programs in biomedicine*, vol. 184, p. 105286, 2020.

[74] H. Makimoto, M. Höckmann, T. Lin, D. Glöckner, S. Gerguri, L. Clasen, J. Schmidt, A. Assadi-Schmidt, A. Bejinariu, P. Müller *et al.*, "Performance of a convolutional neural network derived from an ecg database in recognizing myocardial infarction," *Scientific reports*, vol. 10, no. 1, pp. 1–9, 2020.

[75] M. B. Abubaker and B. Babayiğit, "Detection of cardiovascular diseases in ecg images using machine learning and deep learning methods," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 2, pp. 373–382, 2022.

[76] V. Gliner, N. Keidar, V. Makarov, A. I. Avetisyan, A. Schuster, and Y. Yaniv, "Automatic classification of healthy and disease conditions from images or digital standard 12-lead electrocardiograms," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[77] V. Sangha, B. J. Mortazavi, A. D. Haimovich, A. H. Ribeiro, C. A. Brandt, D. L. Jacoby, W. L. Schulz, H. M. Krumholz, A. L. P. Ribeiro, and R. Khera, "Automated multilabel diagnosis on electrocardiographic images and signals," *Nature communications*, vol. 13, no. 1, pp. 1–12, 2022.

[78] G. A. Tadesse, K. Weldemariam, H. Javed, Y. Liu, J. Liu, J. Chen, and T. Zhu, "Discriminant knowledge extraction from electrocardiograms for automated diagnosis of myocardial infarction," in *Knowledge Management and Acquisition for Intelligent Systems: 17th Pacific Rim Knowledge Acquisition Workshop, PKAW 2020, Yokohama, Japan, January 7–8, 2021, Proceedings 17*. Springer, 2021, pp. 70–82.

[79] Y. Cao, T. Wei, B. Zhang, N. Lin, J. J. Rodrigues, J. Li, and D. Zhang, "Ml-net: Multi-channel lightweight network for detecting myocardial infarction," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 10, pp. 3721–3731, 2021.

[80] E. C. for Practice Guidelines (CPG), J. J. B. C. C. Netherlands), H. B. (Germany), C. C. (Italy), V. D. (France), C. D. (UK), R. F. (Belgium), C. F.-B. (France), D. H. (Israel), A. H. T. Netherlands) *et al.*, "Third universal definition of myocardial infarction," *Journal of the American College of Cardiology*, vol. 60, no. 16, pp. 1581–1598, 2012.

[81] K. Thygesen, J. S. Alpert, A. S. Jaffe, B. R. Chaitman, J. J. Bax, D. A. Morrow, and H. D. White, "Fourth universal definition of myocardial infarction (2018)," *European heart journal*, vol. 40, no. 3, pp. 237–269, 2019.

[82] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[83] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[84] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical image analysis*, vol. 53, pp. 197–207, 2019.

[85] J. R. Hershey and P. A. Olsen, "Approximating the kullback leibler divergence between gaussian mixture models," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 4. IEEE, 2007, pp. IV–317.

[86] C. Shu, Y. Liu, J. Gao, Z. Yan, and C. Shen, "Channel-wise knowledge distillation for dense prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5311–5320.

[87] B. Heo, J. Kim, S. Yun, H. Park, N. Kwak, and J. Y. Choi, "A comprehensive overhaul of feature distillation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1921–1930.

[88] A. Kalyakulina, I. Yusipov, V. Moskalenko, A. Nikolskiy, K. Kosonogov, N. Zolotykh, and M. Ivanchenko, "Lobachevsky university electrocardiography database," *Type: Dataset. Available online: https://physionet.org/content/ludb/1.0. 0/(accessed on 10 July 2021)*, 2020.

[89] P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, "A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg," in *Computers in cardiology 1997*. IEEE, 1997, pp. 673–676.

[90] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[91] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "Ptb-xl, a large publicly available electrocardiography dataset," *Scientific data*, vol. 7, no. 1, p. 154, 2020.

[92] H. Liu, D. Chen, D. Chen, X. Zhang, H. Li, L. Bian, M. Shu, and Y. Wang, "A large-scale multi-label 12-lead electrocardiogram database with standardized diagnostic statements," *Scientific Data*, vol. 9, no. 1, p. 272, 2022.

[93] R. M. Rangayyan, *Biomedical signal analysis*. John Wiley & Sons, 2015.

[94] B. H. Tracey and E. L. Miller, "Nonlocal means denoising of ecg signals," *IEEE transactions on biomedical engineering*, vol. 59, no. 9, pp. 2383–2386, 2012.

[95] C. Liu, X. Zhang, L. Zhao, F. Liu, X. Chen, Y. Yao, and J. Li, "Signal quality assessment and lightweight qrs detection for wearable ecg smartvest system," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1363–1374, 2018.

[96] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[97] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine," *Clinical chemistry*, vol. 39, no. 4, pp. 561–577, 1993.

[98] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 233–240.

[99] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.

[100] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Communications of the ACM*, vol. 64, no. 3, pp. 107–115, 2021.

[101] W. Liu, M. Zhang, Y. Zhang, Y. Liao, Q. Huang, S. Chang, H. Wang, and J. He, "Real-time multilead convolutional neural network for myocardial infarction detection," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1434–1444, 2017.

[102] C. Chen, Y. Wang, J. Niu, X. Liu, Q. Li, and X. Gong, "Domain knowledge powered deep learning for breast cancer diagnosis based on contrast-enhanced ultrasound videos," *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2439–2451, 2021.

[103] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[104] L. Bertinetto, R. Mueller, K. Tertikas, S. Samangooei, and N. A. Lord, "Making better mistakes: Leveraging class hierarchies with deep networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 506–12 515.

[105] T. Yu, S. Kumar, A. Gupta, S. Levine, K. Hausman, and C. Finn, "Gradient surgery for multi-task learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 5824–5836, 2020.

[106] N. Strodthoff, T. Mehari, C. Nagel, P. J. Aston, A. Sundar, C. Graff, J. K. Kanters, W. Haverkamp, O. Dössel, A. Loewe *et al.*, "Ptb-xl+, a comprehensive electrocardiographic feature dataset," *Scientific data*, vol. 10, no. 1, p. 279, 2023.

[107] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet," 1995.

[108] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He *et al.*, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368–1373, 2018.

[109] E. A. P. Alday, A. Gu, A. J. Shah, C. Robichaux, A.-K. I. Wong, C. Liu, F. Liu, A. B. Rad, A. Elola, S. Seyedi *et al.*, "Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020," *Physiological measurement*, vol. 41, no. 12, p. 124003, 2020.

[110] R. M. Rangayyan and S. Krishnan, *Biomedical signal analysis*. John Wiley & Sons, 2024.

[111] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.

[112] Y. Qiang, X. Dong, and Y. Yang, "Automatic detection and localisation of myocardial infarction using multi-channel dense attention neural network," *Biomedical Signal Processing and Control*, vol. 89, p. 105766, 2024.

[113] J. Zhang, M. Liu, P. Xiong, H. Du, H. Zhang, G. Sun, Z. Hou, and X. Liu, "Automated localization of myocardial infarction of image-based multilead ecg tensor with tucker2 decomposition," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–15, 2021.

[114] M. S. Manikandan and S. Dandapat, "Wavelet energy based diagnostic distortion measure for ecg," *Biomedical Signal Processing and Control*, vol. 2, no. 2, pp. 80–96, 2007.

[115] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[116] A. Kosmopoulos, I. Partalas, E. Gaussier, G. Paliouras, and I. Androutsopoulos, "Evaluation measures for hierarchical classification: a unified view and novel approaches," *Data Mining and Knowledge Discovery*, vol. 29, pp. 820–865, 2015.

[117] O. Dekel, J. Keshet, and Y. Singer, "Large margin hierarchical classification," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 27.

[118] E. Begoli, T. Bhattacharya, and D. Kusnezov, "The need for uncertainty quantification in machine-assisted medical decision making," *Nature Machine Intelligence*, vol. 1, no. 1, pp. 20–23, 2019.

[119] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl, "Leveraging uncertainty information from deep neural networks for disease detection," *Scientific reports*, vol. 7, no. 1, pp. 1–14, 2017.

[120] B. Yu, Y. Liu, X. Wu, J. Ren, and Z. Zhao, "Trustworthy diagnosis of electrocardiography signals based on out-of-distribution detection," *PloS one*, vol. 20, no. 2, p. e0317900, 2025.

[121] E. K. Yapp, X. Li, W. F. Lu, and P. S. Tan, "Comparison of base classifiers for multi-label learning," *Neurocomputing*, vol. 394, pp. 51–60, 2020.

[122] S. Nahak and G. Saha, "Ensembled feature based multi-label ecg arrhythmia classification," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2023, pp. 1–4.

[123] D. P. Kingma, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[124] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[125] L. Yang, W. Fan, and N. Bouguila, "Clustering analysis via deep generative models with mixture models," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 1, pp. 340–350, 2020.

[126] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 976–11 986.

[127] D. Hendrycks, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[128] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7482–7491.

[129] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[130] Y. Qiang, X. Dong, X. Liu, Y. Yang, Y. Fang, and J. Dou, "Conv-rgnn: An efficient convolutional residual graph neural network for ecg classification," *Computer Methods and Programs in Biomedicine*, vol. 257, p. 108406, 2024.

[131] Z. Zhuang, M. Liu, A. Cutkosky, and F. Orabona, "Understanding adamw through proximal methods and scale-freeness," *arXiv preprint arXiv:2202.00089*, 2022.

[132] A. Lyon, A. Mincholé, J. P. Martínez, P. Laguna, and B. Rodriguez, "Computational techniques for ecg analysis and interpretation in light of their contribution to medical advances," *Journal of The Royal Society Interface*, vol. 15, no. 138, p. 20170821, 2018.

[133] R. Sassi, R. R. Bond, A. Cairns, D. D. Finlay, D. Guldenring, G. Libretti, L. Isola, M. Vaglio, R. Poeta, M. Campana *et al.*, "Pdf–ecg in clinical practice: A model for long–term preservation of digital 12–lead ecg data," *Journal of electrocardiology*, vol. 50, no. 6, pp. 776–780, 2017.

[134] S. Bhaskarpandit, A. Gade, S. Dash, D. K. Dash, R. K. Tripathy, and R. B. Pachori, "Detection of myocardial infarction from 12-lead ecg trace images using eigendomain deep representation learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023.

[135] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue, "Arbitrary-oriented scene text detection via rotation proposals," *IEEE transactions on multimedia*, vol. 20, no. 11, pp. 3111–3122, 2018.

[136] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 9268–9277.

[137] R. Girshick, “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[138] Y. Liu, C. Qin, C. Liu, J. Liu, Y. Jin, Z. Li, and L. Zhao, “Multiple high-regional-incidence cardiac disease diagnosis with deep learning and its potential to elevate cardiologist performance,” *Iscience*, vol. 25, no. 11, 2022.

[139] P. N. Casale, R. B. Devereux, P. Kligfield, R. R. Eisenberg, D. H. Miller, B. S. Chaudhary, and M. C. Phillips, “Electrocardiographic detection of left ventricular hypertrophy: development and prospective validation of improved criteria,” *Journal of the American College of Cardiology*, vol. 6, no. 3, pp. 572–580, 1985.

[140] S. Bhattarai, L. Chhabra, M. F. Hashmi, and C. Willoughby, “Anteroseptal myocardial infarction,” 2019.

[141] Z. Nesheiwat, A. Goyal, M. Jagtap, and A. Shammas, “Atrial fibrillation (nursing),” in *StatPearls [Internet]*. StatPearls Publishing, 2021.

[142] M. Refaat, M. Mansour, J. P. Singh, J. Ruskin, and E. K. Heist, “Electrocardiographic characteristics in right ventricular vs biventricular pacing in patients with paced right bundle-branch block qrs pattern,” *Journal of Electrocardiology*, vol. 44, no. 2, pp. 289–295, 2011.

[143] D. G. Strauss, R. H. Selvester, and G. S. Wagner, “Defining left bundle branch block in the era of cardiac resynchronization therapy,” *The American journal of cardiology*, vol. 107, no. 6, pp. 927–934, 2011.

LIST OF PUBLICATIONS

Journal Publications:

1. **P. S. Choudhary** and S. Dandapat, “Morphology-aware ecg diagnostic framework with cross-task attention transfer for improved myocardial infarction diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, 2024.

2. **P. S. Choudhary**, L. N. Sharma and S. Dandapat, ”Joint Detection of Rhythmic and Morphological Abnormalities in Electrocardiographic Images: A Multitask Learning Approach,” *IEEE Transactions on Artificial Intelligence*, 2025.

3. **P. S. Choudhary** and S. Dandapat, “Multi-Task Physics-Constrained Approach for Hierarchical Classification of Myocardial Infarction from Electrocardiogram Signal,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **(Revision submitted)**.

4. **P. S. Choudhary** and S. Dandapat, “Uncertainty-Aware Multi-Label Electrocardiogram diagnostic framework”, **(manuscript under preparation)**.

Conference Publications:

1. **P. S. Choudhary** and S. Dandapat, ”An evaluation of machine learning classifiers for detection of myocardial infarction using wavelet entropy and eigenspace features.” *IEEE applied signal processing conference (ASPCON)*, pp. 222-226. IEEE, 2020.

2. R.K.S Sane, **P. S. Choudhary**, L. N. Sharma and S. Dandapat, ”Detection of myocardial infarction from 12 lead ECG images.” *National conference on communications (NCC)*, pp. 1-6. IEEE, 2021.

3. **P. S. Choudhary** and S. Dandapat, ”Multibranch 1D CNN for detection and localization of myocardial infarction from 12 lead electrocardiogram signal.” *IEEE 18th India Council International Conference (INDICON)*, pp. 1-5. IEEE, 2021.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati