Automated Diagnosis of Cardiac Disorders from Electrocardiogram Signals using Deep Learning

Eedara Prabhakararao

Automated Diagnosis of Cardiac Disorders from Electrocardiogram Signals using Deep Learning

A

Thesis submitted

for the award of the degree of

DOCTOR OF PHILOSOPHY

By

Eedara Prabhakararao

DEPARTMENT OF ELECTRONICS AND ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781 039, ASSAM, INDIA

MARCH 2023

Certificate

This is to certify that the thesis entitled "**Automated Diagnosis of Cardiac Disorders from Electrocardiogram Signals using Deep Learning**", submitted by **Eedara Prabhakararao** (166102109), 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 my 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.

Prof. Samarendra Dandapat,
Professor
Department of Electronics and Electrical Engineering
Indian Institute of Technology Guwahati
Guwahati - 781 039, Assam, India.

Dated: 23-09-2022
Guwahati.

To

The Almighty

for his blessings

My beloved parents

Mohanrao Eedara and Yesoda Eedara

for their blessings, love and encouragement

My guide

Prof. Samarendra Dandapat

for his guidance, support and inspiration

&

My sister and brother

Raji and Ganesh

for their love and support

Acknowledgements

First of all, I would like to thank God. He has given me strength, knowledge, and encouragement throughout all the challenging moments of completing this thesis. I am truly grateful for His unconditional and endless love, mercy, and grace. This thesis would not have been possible without the encouragement and support received from many people. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this thesis.

I would like to express my sincere and utmost gratitude to my research supervisor, Prof. Samarendra Dandapat, for believing in me and giving me the opportunity to be part of the Electro-Medical and Speech Technology (EMST) research group at the IIT Guwahati. Prof. Dandapat taught me how to do the research and gave me the freedom and the support to pursue research ideas while providing me with essential guidance towards achieving the research goals. I am indebted to him for his invaluable time, help, guidance, insightful discussions, and enormous moral support throughout my research work.

I am very much thankful to Prof. P. K. Bora, the chairman of my doctoral committee, for his constant support and encouragement throughout my Ph.D. period. I am grateful to the other members of my doctoral committee, Dr. Tony Jacob, Dr. S. Kashyap, and Dr. R. Kulkarni, for the insightful comments and constructive criticisms, which helped me bring my research work to the current form. I also would like to thank all the anonymous reviewers for their valuable suggestions about my research articles, which helped improve the technical quality of my dissertation. I would also like to thank all other faculty members of the Department of Electronics and Electrical Engineering, IIT Guwahati, for their care and support. I am also very thankful to all the technical, office, security, canteen, and maintenance staff members of the Department for their help when required. I sincerely thank my M.Tech. supervisor, Dr. M. S. Manikandan, for introducing me to the exciting field of biomedical signal processing at IIT Bhubaneswar.

I thank the PhysioNet repository and the MIT Laboratory for Computational Physiology members for publicly releasing crucial clinical ECG datasets that were used to carry out this research work.

I am very much thankful to my senior members of IIT Guwahati for their uplifting conversations, love, support, and care, which helped me grow as a better person. I am also thankful to my dear friends and lab mates of the EMST laboratory for their help and moral support during my Ph.D. journey. Above all, I sincerely thank all my family members for their endless love, support, and influence in my life. Without their support, love, and sacrifice, I would have never gotten this far. Thank you for believing in me.

Eedara Prabhakararao

Abstract

Cardiac disorders are the leading causes of human mortality worldwide. Most disorders progressively worsen over time; as a result, if left untreated can lead to severe complications such as heart attack and stroke. Therefore, early diagnosis and a better understanding of the disease progression are crucial to timely initiating appropriate treatment plans that can help prevent further disease progression and severe cardiac events. Widely used by clinicians as a routine modality in hospitals, electrocardiogram (ECG) recordings non-invasively capture the heart's electrical activity from the body surface. Many cardiac electrophysiological abnormalities have a signature on the ECG, and their identification can aid in the early diagnosis of cardiac disorders. In practice, an experienced cardiologist manually examines the 12-lead ECGs or the single-lead Holter ECG to diagnose disorders. Manual examination of enormous amounts of data can be tedious, time-consuming, and prone to human errors. Hence, research on automated ECG interpretation methods has been gaining popularity as they can aid in rapid and improved objective clinical decision-making, allowing clinicians to provide timely treatment. In order to develop effective automated diagnosis methods, several key challenges must be addressed, including wide morphological and temporal association variabilities in pathological ECG characteristics across cardiac abnormalities, subtle ECG changes within the same cardiac disorder based on severity, intermittent nature of rhythm disorders, diagnostic redundancy of 12-lead ECG, and susceptibility of ECG signals for different noises. This dissertation focuses on developing automated methods that can address the above challenges for the efficient and reliable diagnosis of various cardiac disorders from single- and 12-lead ECG signals.

Myocardial infarction (MI) is a lethal heart condition that occurs due to the blockage of blood flow to the heart muscle. Although MI is a progressive disorder, existing methods consider it as one broad category for the diagnosis and do not address its severity. In practice, the severity assessment of MI in terms of early MI, acute MI, and chronic MI using 12-lead ECGs can play a crucial role in timely initiating appropriate treatment to prevent further disease progression and reduce mortality. However, subtle variabilities in the pathological ECG manifestations and dynamic changes in their temporal dependencies during the disease progression pose a challenge for reliable severity staging. In addition, the implicit diagnostic redundancy of the

12-lead ECG obscures the subtle pathological manifestations, degrading diagnostic accuracy. The first part of the thesis presents a multi-lead diagnostic attention-based recurrent neural network (MLDA-RNN) method that can address these challenges and extract discriminative feature representations from 12-lead ECG to improve the MI severity stages classification. The method systematically processes 12-lead ECG using lead-specific RNN layers followed by intra- and inter-lead attention modules. The RNNs and attention modules effectively characterize the temporal dynamics of subtle pathological manifestations and emphasize diagnostically relevant 12-lead ECG features to obtain high-level discriminative representations for classification. The experimental results show that the MLDA-RNN method achieves improved MI severity stages classification. Additionally, the qualitative results reveal that the MLDA-RNN automatically selects salient ECG leads and focuses on clinically relevant ECG features during the diagnosis, thereby mimicking the cardiologist's way of clinical 12-lead ECG interpretation.

Congestive heart failure (CHF) is a complex clinical syndrome; thus, it results in wide variability in the pathological ECG features and their temporal dependencies. Existing methods have mainly used deep convolutional neural networks (DCNNs) to diagnose CHF. These deeper networks are hard to optimize because of vanishing gradients, difficult to interpret, and show limited detection accuracy as they fail to effectively capture the complex pathological ECG features and their temporal dependencies associated with CHF syndrome. The second part of the thesis presents a hybrid model named attention-based deep residual RNN (A-DRRNet) to address the above challenges for accurate CHF diagnosis from single-lead ECG. The model consists of multi-layered RNNs with residual connections followed by an attention module to attentively encode temporal dependencies (P-QRS-T) of complex pathological ECG features by maintaining proper gradients at the earlier layers for efficient CHF detection. The experimental results verify that the proposed method has superior diagnostic accuracy, light-weight architecture, and promising model transparency visualizations compared to the existing methods.

Atrial fibrillation (AF) is the most common heart rhythm disorder that can progressively worsen over time. AF can be intermittent and asymptomatic; as a result, it often remains undiagnosed until it causes a life-threatening condition, such as stroke and thromboembolism. It is thus a disorder that would benefit from long-term continuous monitoring with Holter ECG devices. Analyzing AF burden (the percentage of time an individual is in AF rhythm) from the long-term recordings can provide improved prognostic value compared to the traditional binary AF diagnosis (present or absent) methods using snapshot ECG. However, the presence of frequent ectopic beats and different noise levels pose a challenge for precise AF burden estimation. The third part of the thesis proposes a multi-task DCNN (MT-DCNN) method for accurately estimating AF burden from the long-term ECG recordings. The model consists of AF detection as a primary task and reconstruction of ECG sequence as an auxiliary task using DCNNs. The auxiliary task regularizes the model to learn robust feature representations for efficient AF detection, thereby aiding accurate AF burden estimation. The experimental results on four diverse Holter ECG datasets with frequent ectopic beats and different noise levels verify that the proposed framework provides a lesser mean AF burden estimation error over the baseline approaches. In addition, the analysis of daily AF burden suggests that it effectively improves the diagnosis and stroke risk stratification in AF patients; thus, it has the potential to be used in remote patient monitoring applications for personalized AF diagnostics and management.

Apart from disease-wise analysis, the automated methods should also be able to analyze 12-lead ECGs of multiple cardiac disorders such as MI, hypertrophic cardiomyopathy (abnormal thickening of heart muscle), and conduction disturbances (electrical conduction system dysfunction) simultaneously and identify the underlying disorder for its appropriate management. However, these diseases show subtle and similar pathological ECG changes with wide morphological variabilities (multi-scale), challenging their reliable classification. In practice, clinicians often seek advice from several experts for patients with confusing ECG biomarkers before making the final diagnosis. Inspired by this, the last part of the thesis presents a multi-scale deep temporal CNN ensemble (MS-DTCE) framework that can extract and combines diverse multi-scaled ECG features for effectively classifying multiple cardiac disorders. The MS-DTCE consists of several temporal CNN-based expert classifiers designed using dilated and causal convolutional filters with different receptive fields to extract the scale-specific features from the 12-lead ECG and generate local predictions. The MS-DTCE also includes a temporal CNN-based gating network that aggregates the local predictions of experts based on their competencies and generates the final diagnosis decisions. The experimental results on the clinical 12-lead ECG datasets show that the MS-DTCE framework significantly improves the classification performance for multiple cardiac ailments over existing methods.

Keywords: Electrocardiogram, Cardiac disorders, Classification, Deep learning, Recurrent neural networks, Convolutional neural networks, Attention mechanism.

Contents

List of Figures

List of Tables

List of Acronyms

AB	Atrial Bigeminy
Acc	Accuracy
A-DRRNet	Attention-based Deep Residual Recurrent Neural Network
AF	Atrial Fibrillation
AFDB	Atrial Fibrillation Database
AFL	Atrial Flutter
AHA	American Heart Association
ALAR	Across Lead Attentive Representation
AMI	Acute MI
ANN	Artificial Neural Network
ANS	Autonomic Nervous System
ASMI	Anterior Septal Myocardial Infarction
ATCNN	Attention-based temporal Convolutional Neural Network
ATI-CNN	Attention-based Time Incremental Convolutional Neural Network
AUC	Area Under Curve
AV	Atrio-Ventricular
AVNN	Average of Normal-to-Normal Intervals
BAR	Beat Attentive Representation
BBB	Bundle Branch Block
BCE	Binary Cross Entropy
BCI	Brain Computer Interface
BIDMC	Beth Israel Deaconess Medical Centre
BiGRU	Bidirectional Gated Recurrent Unit
BiLSTM	Bidirectional Long Short Term Memory
BN	Batch Normalization

BPF	Band Pass Filter
CABG	Coronary Artery Bypass Graft
CD	Conduction Disturbance
CDAE	Convolutional Denoising Auto-Encoder
CF	Clinical Features
CHF	Congestive Heart Failure
CMI	Chronic Myocardial Infarction
CNN	Convolutional Neural Network
CosEn	Cosine Entropy
CPSC	China Physiological Signal Challenge
CT	Computarized Tomography
CV	Cross-Validation
CVD	Cardiovascular Disease
DbD	Distance based Disagreement
DCNN	Deep Convolutional Neural Network
DF	Dominant Frequency
DFT	Discrete Fourier Transform
DL	Deep Learning
DMSFNet	Deep Multi-Scale Fusion Network
DRNN	Deep Recurrent Neural Network
DTCNN	deep Temporal Convolutional Neural Network
ECG	Electrocardiogram
ECHO	Echocardiography
EMD	Emperical Mode Decomposition
EMI	Early Myocardial Infarction
EWT	Emperical Wavelet Transform
FC	Fully Connected
FCNN	Fully Convolutional Neural Network
FN	False Negative
FP	False Positive
fWP	f-Wave Power

fWPMAW	f-Wave Power in Main Atrial Wave
GAN	Generative Adversarial Network
GAP	Global Average Pooling
GRACE	Global Registry of Acute Coronary Events
Grad-CAM	Gradient-weighted Class Activation Mapping
GRU	Gated Recurrent Unit
HC	Healthy Control
HF	High Frequency
HOSVD	Higher Order Singular Value Decomposition
HPF	High Pass Filter
HRV	Heart Rate Variability
HYP	Hypertrophy
I-AVB	First-degree Atrio-Ventricular Block
IMI	Inferior Myocardial Infarction
IVR	Idio-Ventricular Rhythm
KNN	K-Nearest Neighbours
LAD	Left Anterior Descending
LBBB	Left Bundle Branch Block
LCx	Left Circumflex
LF	Low Frequency
LMCA	Left Main Coronary Artery
LPF	Low Pass Filter
LS-TCNN	Lead Specific Temporal Convolutional Neural Network
LSTM	Long Short-Term Memory
LTAFDB	Long-Term Atrial Fibrillation Database
LTNSRDB	Long-Term Normal Sinus Rhythm Database
LVEF	Left Ventricular Ejection Fraction
LVH	Left Ventricular Hypertrophy
meaRR	Mean of RR-intervals
MFB-CNN	Multiple Feature Branch Convolutional Neural Network
MI	Myocardial Infarction

ML	Machine Learning
MLDA-RNN	Multi Lead Diagnostic Attention-based Recurrent Neural Network
MLP	Multi-Layer Perceptron
ML-ResNet	Multi Lead Residual Network
M-MI	Mimicking Myocardial Infarction
MoE	Mixture of Experts
MPI	Myocardial Perfusion Imaging
MRI	Magnetic Resonance Imaging
MS-CNN	Multi-Scale Convolutional Neural Network
MS-DTCE	Multi-Scale Deep Temporal Convolutional Neural Network Ensemble
MSE	Mean Squared Error
MT-DCNN	Multi-Task Deep Convolutional Neural Network
nMcc	Normalized Matthews Correlation Coefficient
NMT	Neural Machine Translation
NSR	Normal Sinus Rhythm
NSRDB	Normal Sinus Rhythm Database
NYHA	New York Heart Association
OA	Overall Accuracy
OS	Overall Sensitivity
OSp	Overall Specificity
OTH	Others
PAC	Premature Atrial Contraction
PAF	Paroxysmal Atrial Fibrillation
PCI	Percutaneous Coronary Intervention
pNN20	Percentage of successive NN intervals that differ by more than 20 ms
pNN50	Percentage of successive NN intervals that differ by more than 50 ms
PPV	Positive Predictive Value
Pr	Precision
PTB	Physikalisch-Technische Bundesanstalt
PVC	Premature Ventricular Contraction
QoL	Quality of Life

RBBB	Right Bundle Branch Block
RCA	Right Coronary Artery
Re	Recall
ReLU	Rectified Linear activation Unit
RF	Random Forest
RMSSD	Root Mean Square of the Successive Differences
RNN	Recurrent Neural Network
SA	Sino-Atrial
SampEnfW	Sample Entropy of f-Wave signal
SBR	Sinus Bradycardia
SCN	Spectral Concentration
SD	Standard Deviation
SD-DTCNN	Scale Dependent Deep Temporal Convolutional Neural Network
SDNN	Standard deviation of NN intervals
Se	Sensitivity
SNR	Signal to Noise Ratio
Sp	Specificity
STD	ST-segment Depression
STE	ST-segment Elevation
SVM	Support Vector Machines
SVTA	Supra-Ventricular Tachyarrhythmia
TA-CNN	Temporal Attention-based Convolutional Neural Network
TCNN	Temporal Convolutional Neural Network
TIMI	Thrombolysis in Myocardial Infarction
TN	True Negative
TP	True Positive
VMD	Variational Mode Decomposition
WLAR	Within Lead Attentive Representation
WT	Wavelet Transform

1

Introduction

Contents

Cardiovascular diseases (CVDs), a group of heart and blood vessel disorders, are the leading cause of morbidity and mortality worldwide [1,2]. The number of people dying from CVDs is rising at an alarming rate, including 18.6 million deaths in 2019, representing 32% of global deaths [3]. The majority of CVD deaths are attributed to cardiac disorders, including coronary artery disease or myocardial infarction (MI), hypertrophic cardiomyopathy (HYP), conduction disturbances (CD), atrial fibrillation (AF), and congestive heart failure (CHF), and so on [4]. Often, these cardiac disorders progressively worsen over time; thus, if left untreated can lead to severe complications such as heart attack and stroke [5]. Therefore, early diagnosis and a better understanding of the disease progression are crucial to the timely initiation of appropriate treatment strategies that could help improve outcomes in patients with cardiac disorders [6].

Current diagnostic tools for the assessment of cardiovascular health include electrocardiography (ECG), echocardiography (ECHO), computerized tomography (CT) calcium scoring, CT coronary angiography, cardiac magnetic resonance imaging (MRI), coronary catheter angiography, and myocardial perfusion imaging (MPI) [7]. Among all, ECG is a simple, rapid, non-invasive, and inexpensive tool widely used by cardiologists and medical practitioners for recording the electrical activity of the heart from the body's surface [8]. Therefore, many cardiac structural or electrophysiological abnormalities have a signature on the ECG, and their accurate identification can aid in the diagnosis of cardiac disorders [8,9]. In a clinical context, ECG recordings are acquired in two different ways: (i) the standard 12-lead ECGs provide comprehensive information on cardiac activity from 12 different perspectives (leads) over several heartbeats and are typically used to diagnose spatially occurring acute cardiac disorders such as MI, CD, and HYP; (ii) on the other hand, single-lead ambulatory or Holter ECGs record the electrical activity of the heart from only one perspective over more extended periods (several hours), which aid in the early diagnosis and characterization of the paroxysmal rhythm abnormalities such as AF [9]. Therefore, careful analysis of cardiac electrical activity using single- or 12-lead ECG signals is of great importance for the diagnosis, risk assessment of cardiac disorders, and understanding of the effect of clinical interventions [10].

In clinical practice, an experienced cardiologist manually examines the changes in typical ECG characteristics such as P-wave, QRS-complex, ST-segment, and T-wave to diagnose cardiac disorders [11]. Due to the diverse etiology of cardiac disorders, the pathological ECG waveform characteristics dynamically vary in terms of amplitude, duration, morphology, and temporal association [8,11]. Moreover, based on disease progression, the pathological ECG features undergo subtle variations at specific leads of 12-lead ECG [10]. Manual analysis of 12-lead or long-term single-lead ECGs to diagnose cardiac ailments can be tedious, time-consuming, and prone to errors [12]. Even for an experienced clinician, accurately interpreting the subtle ECG changes during disease progression is challenging [13]. Also, visual inspection of the diverse pathological manifestations is subjective, causing inter- and intra-rater variability [12, 13]. Consequently, there is an urgent unmet need for automated ECG interpretation methods as they can effectively represent the disease variabilities and aid in rapid and improved objective clinical decision-making.

Furthermore, in recent years, the widespread digitization of ECG data and the advances in sensor and wireless technologies have led to the development of various new portable and wearable ECG acquisition devices [14]. The rise of these novel technologies has resulted in the acquisition of more than 300 million ECGs annually worldwide [15]. Manual analysis of these enormous amounts of data is often challenging because experienced medical practitioners are scarce [16]. In addition, the ECG recordings acquired from portable or wearable devices are prone to various noises, which degrades ECG signal quality and affects its clinical utility [9, 14]. Therefore, the advancement in ECG acquisition technologies and the high prevalence of CVDs in the present population have intensified the research towards developing automated diagnosis systems. These systems can play a crucial role in the early diagnosis of cardiac disorders and have the potential to assist clinicians in making improved diagnosis decisions [12, 13].

This dissertation focus on developing efficient and reliable automated methods for the diagnosis of various cardiac disorders using single- and 12-lead ECG signals. The chapter begins with a brief introduction to ECG waveform and its pathological manifestations during cardiac abnormalities, followed by a review of the existing automated methods for ECG-based cardiac disorders diagnosis. We then highlight the motivation of the research work. Following that, we present the contributions and outline of the thesis.

1.1 The Electrocardiogram Signal

As a non-invasive recording of surface potentials induced by cardiac electrical activity, ECG provides critical information about the heart's functioning and helps assess cardiac health [8]. In the following section, we briefly discuss the main concepts related to the physiological and technical basis of the ECG signal.

1.1.1 Physiological Origin

The human heart is a rhythmically beating muscular organ responsible for pumping oxygen and nutrients rich blood throughout the body. The perfect rhythmic functioning of the heart is coordinated by two types of cardiac muscle cells, i.e., contractile cells and specialized muscle cells [17]. Contractile cells provide the powerful contractions that propel the blood, while the specialized muscle cells of the conducting system coordinate these contractions. The electrical conduction system of the heart or pathway consists of a sino-atrial (SA) node, atrio-ventricular (AV) node, the bundle of His, bundle branches (left and right), and Purkinje fibers, as depicted in Figure 1.1. These conduction system components enable the synchronized

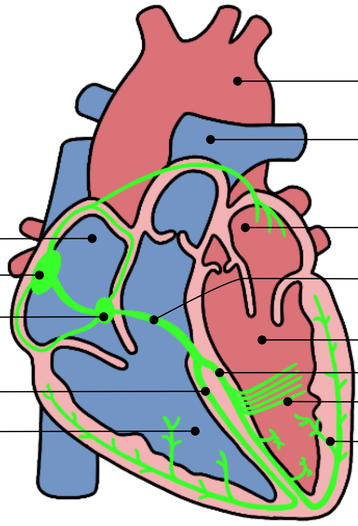


Figure 1.1: Electrical conduction system of the heart. The figure illustrates the sequential activation of different electrical nodes and muscle fibers of the normal electrical conduction system pathway.

contraction and relaxation of the contractile muscle cells, which generate an electrical potential difference that can be measured by placing electrodes on the body surface [17]. The resulting amplified signal is known as the ECG [8, 17]. As shown in Figure 1.2, each wave or segment of the ECG corresponds to important electrical events in the heart. The following steps briefly describe the relationship between these events and the ECG characteristics [17].

- Normal cardiac rhythm is established by the spontaneous depolarization of the SA node, also known as the natural pacemaker of the heart. The spread of this depolarization electrical impulse corresponds with the iso-electrical line preceding the P-wave.
- The electrical impulse is then spread throughout the atria via specialized pathways and reaches the AV node within 50 ms. Along the way, the conducting cells stimulate the myocardial contracting cells of both the atria, making them depolarize or contract, resulting in a local P-wave in the ECG signal.
- The AV node conducting cells are smaller in size, which slows down the electrical impulse by 100 ms before reaching the ventricles. This action leads to the iso-electrical PR-segment in the ECG.
- The electrical impulse then travels through the bundle of His, bundle branches, and Purkinje fibers and rapidly spreads across the ventricles, making them contract or depolarize, resulting in a large QRS-complex in the ECG signal. The passage of ventricular depolarization takes about 100 ms.
- After depolarization, the ventricular muscle cells reach a plateau phase, which lasts about 120 ms, during which no electrical activity occurs. It corresponds to the iso-electric ST-segment in the ECG.

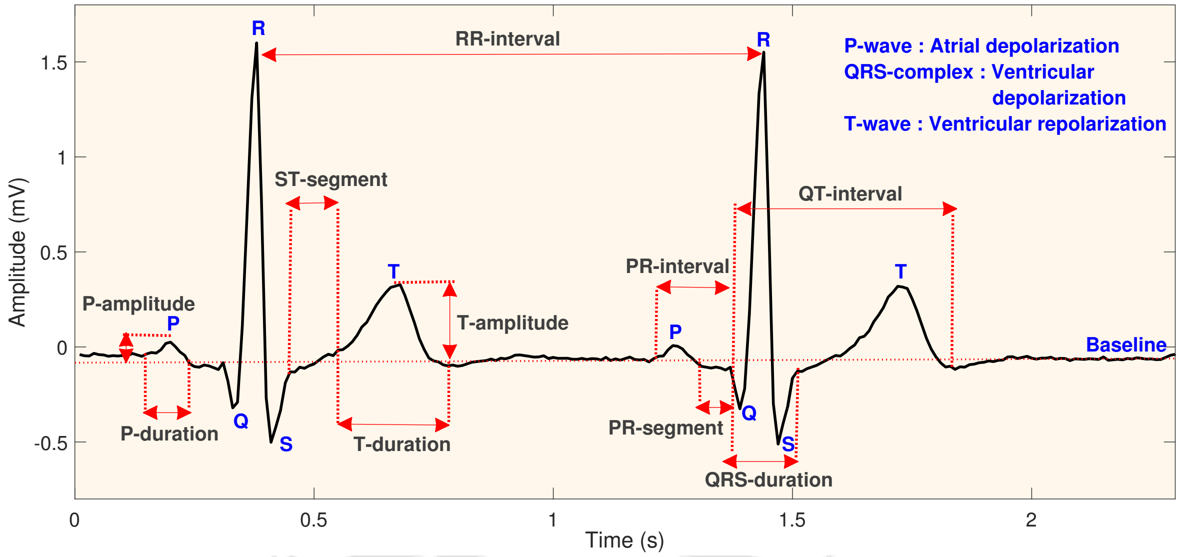


Figure 1.2: Morphological characteristics of normal sinus rhythm ECG. Accurate ECG interpretation requires the knowledge of amplitudes, durations, shapes of the morphological characteristics, and their temporal relationships.

- Finally, the ventricles undergo repolarization, which causes the T-wave in the ECG. The passage of ventricular repolarization takes approximately 160 ms.

The above-described sequential process constitutes a single ECG heartbeat; thus, temporal dependencies naturally exist in this waveform. This cardiac cycle regularly repeats in normal or healthy control (HC) subjects, leading to a stable rhythm in the ECG known as normal sinus rhythm (NSR). It is evident from Figure 1.2 that the ECG contains crucial information about different electrical events in the heart; thus, any abnormality in the heart's functioning can result in visible morphological changes in the ECG signal. Therefore, studying these ECG morphological changes in terms of shape, amplitude, and duration and their temporal relationships provides valuable insights into cardiac pathophysiology [5, 17]. In practice, the popular ECG recording systems for cardiac health assessment include the standard 12-lead ECG and the single-lead ECG [9]. In the following sections, we briefly describe these ECG recording systems.

1.1.2 Standard 12-Lead ECG

In an ECG recording system, the term "lead or channel" refers to the difference in electrical potential between two electrodes. In practice, cardiologists consider the 12-lead ECG configuration as a gold standard tool for diagnosing various cardiac ailments [11]. The 12-lead ECG can be recorded by placing ten electrodes in a specific order on the patient's skin, including four electrodes placed on the limbs and six on the chest. Figure 1.3 (a) illustrates the placement of electrodes for recording 12-lead ECG. The electrodes are labeled according to their location on the body surface (LA: Left Arm, RA: Right Arm, LL: Left Leg, RL:

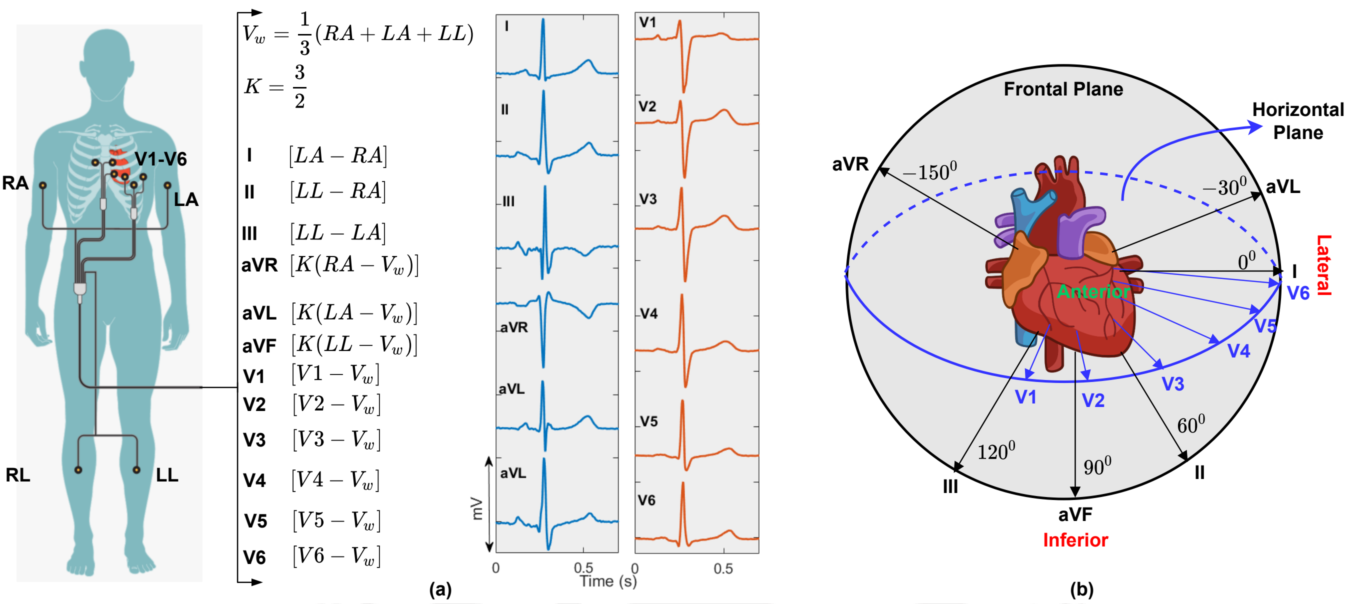


Figure 1.3: The standard 12-lead ECG recording system. (a) Illustrates the standard placement of ten electrodes on the human body for recording 12-lead ECG, including three limb leads (I, II, and III), three augmented limb leads (aVR, aVL, and aVF), and six chest leads (V1-V6). The figure also depicts the electrical potential differences between the electrodes and corresponding 12-lead ECG data. (b) The twelve ECG leads reflect different spatial angles of the heart and gives a comprehensive three-dimensional view of the cardiac electrical behaviour in frontal and horizontal planes. Specifically, leads II, III, and aVF monitor, the inferior heart region, leads V1-V4 monitor the anterior region, and leads I, aVL, V5, and V6 view the lateral region.

Right Leg, and Vx: Chest electrodes V1 to V6). The 12 ECG leads recorded from these electrodes can be separated into limb leads, augmented limb leads, and chest leads. The limb leads are derived from the three limb electrodes, i.e., lead I: voltage difference between LA and RA (LA-RA), lead II: LL-RA, and lead III: LL-LA. Note that the RL electrode is considered to be a reference electrode. The combination of limb leads forms a virtual reference electrode known as Wilson's central terminal (V_w) [11]. The augmented limb leads aVR, aVL, and aVF measure the potential difference between one of the limb electrodes (RA/LA/LL) and a virtual electrode, which provides additional views. Finally, the chest or precordial leads V1 to V6 measures the potential difference between the chest electrodes and the virtual reference electrode (V_w).

Each ECG lead views the heart from a different spatial angle or perspective (see Figure 1.3 (b)). Thus, the twelve ECG leads completely characterize the heart's electrical activity and provide a comprehensive three-dimensional view. Specifically, the limb and the augmented limb leads provide the cardiac electrical activity information in the frontal plane of the body, whereas the chest leads provide information in the horizontal plane (Figure 1.3 (b)). Clinically, the twelve ECG leads can be further categorized into three groups based on which anatomical area of the heart can be monitored with that ECG lead (see Figure 1.3 (b)) [11]. The relationship between the heart region and the ECG leads is as follows: inferior region (leads II, III, and aVF), anterior region (leads V1-V4), and lateral region (leads I, aVF, V5, and V6). Monitoring the functioning of different cardiac regions provides crucial clinical information that allows diagnosis of the localized cardiac abnormalities such as myocardial infarction (anterior, lateral, and inferior), conduction disturbances (left and right bundle branch blocks), and hypertrophic cardiomyopathy (left and right) [10, 11].

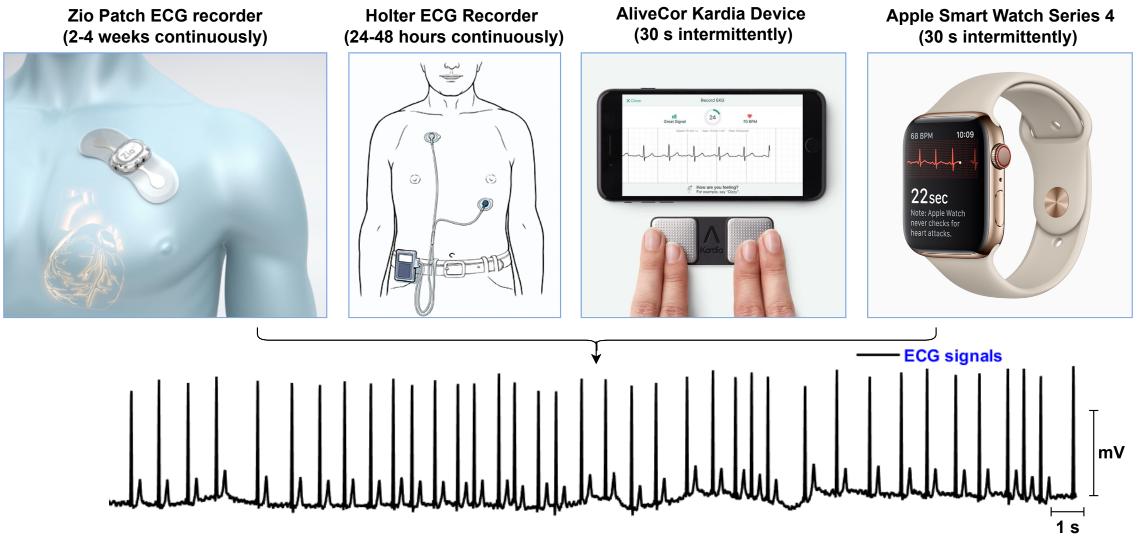


Figure 1.4: Different single-lead ambulatory ECG recording devices with their electrode placement and cardiac monitoring periods.

1.1.3 Single-Lead ECG

The 12-lead ECGs are widely used in hospital settings and are often short-term recordings of several heartbeats. However, when the patient needs to be monitored continuously for several hours, the 12-lead ECG is impractical as the patient must be attached to ten electrodes [9]. Long-term recordings can be acquired using Holter or portable ECG monitors. Clinicians often recommend these Holter monitors to characterize the occasionally occurring abnormal events that have high chances of being missed during the hospital visits [14]. In addition, remote patient monitoring using Holter ECG recorders can play a crucial role in the early diagnosis and management of silent or asymptomatic cardiac disorders. The recent developments in sensor and wireless technologies have led to several new ECG recording devices (Figure 1.4) [14, 18]. Most of these devices record single-lead ECG, typically of lead I or II. While portable Patch and Holter ECG devices are well suited for long-term ambulatory monitoring, handheld mobile and Smartwatch devices are suitable for intermittent monitoring, where patients can take an ECG measurement whenever they experience an abnormal sensation. Although 12-lead ECG is clinically superior, life-threatening rhythm disorders such as AF and CHF can be reliably diagnosed with single-lead ECG recordings [18].

In short, the single-lead provides intra- and inter-heartbeat diagnostic information that helps analyze various heart rhythm abnormalities. On the other hand, the rich diagnostic information of the standard

12-lead ECG, i.e., intra-lead (intra- and inter-beat) and inter-lead information can help diagnose spatially occurring acute cardiac disorders. Therefore, based on the clinical application, 12-lead ECGs or long-term single-lead ECG recordings might be preferred for the analysis.

1.2 Cardiac Disorders and Pathological Changes in ECG Signals

This dissertation mainly focuses on diagnosing acute cardiac disorders such as MI, bundle branch blocks (BBB), HYP, AF, and CHF. The following sections briefly present the pathological variations in ECG due to the aforementioned cardiac disorders.

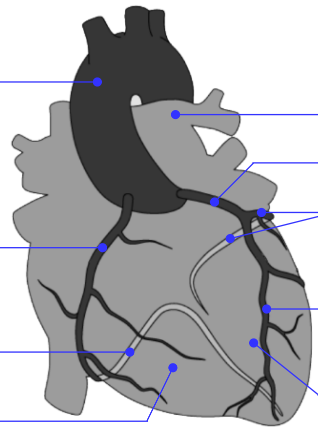


Figure 1.5: Illustrates the anterior view of the heart's coronary arteries and their anatomical relations.

1.2.1 Myocardial Infarction

Myocardial infarction (MI) is a lethal heart condition that involves the reduction of blood supply to the heart tissue or myocardium due to occlusion in the coronary arteries [5]. The coronary arteries are a group of arterial blood vessels that wrap around the heart and transport oxygenated blood to the entire myocardium (see Figure 1.5). The left main coronary artery (LMCA) perfuses or supplies blood to the left side of the heart. It branches into two arteries, the left anterior descending artery (LAD) and the left circumflex artery (LCx), where LAD mainly supplies blood to the anterior wall of the left ventricle and the interventricular septum, and the LCx artery supplies to the lateral wall of the left ventricle. Similarly, the right coronary artery (RCA) supplies oxygenated blood to the right ventricle, SA, and AV node, whereas its posterior descending artery supplies blood to the inferior and posterior walls of the left ventricle [5, 17].

During MI, one or more coronary arteries are occluded with the progressive atherosclerotic plaque, reducing oxygenated blood and nutrients to the myocardium [19]. The reduced oxygen supply causes

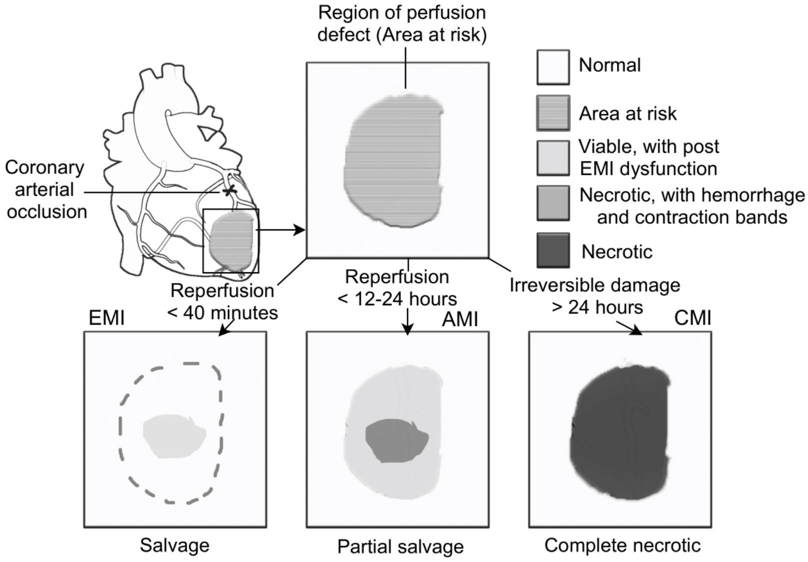


Figure 1.6: Illustrates the consequences of reperfusion, i.e., restoring blood flow through an occluded coronary artery at different severity stages of MI, such as early MI (EMI), acute MI (AMI), and chronic MI (CMI) [5,21].

damage to the myocardial tissues, which slowly expands across the "myocardium at risk" and finally manifests as complete necrosis [20]. The progression of MI severity can be broadly categorized into three stages, such as early MI (EMI), acute MI (AMI), and chronic MI (CMI), based on the time from the onset of the symptoms [5,21]. As illustrated in Figure 1.6, reperfusion of the occluded artery within 40 minutes can salvage the entire myocardium at risk. This stage of MI is reversible and is called EMI [5]. If the occlusion persists for 12-24 hours, cardiac tissue death starts, and it causes irreversible myocardial damage known as AMI [5]. If the occlusion persists without treatment for more than 24 hours, the myocardium at risk becomes completely necrotic. This irreversible damage to the heart muscle is known as CMI [5]. The CMI causes severe ventricular remodeling with more than 90% mortality [21]. Based on the severity of the disease, its management involves using fibrinolytic therapy in the early stages and percutaneous coronary intervention (PCI) or coronary artery bypass graft (CABG) in the acute or chronic stages [21]. Therefore, early diagnosis and risk assessment of MI is crucial for the timely initiation of appropriate treatment procedures, which improves viable myocardium salvage and prognosis [22].

Clinicians widely use the standard 12-lead ECG to diagnose MI [23]. During MI progression, the ECG characteristics in the leads facing the area of infarction undergo subtle changes like T-wave peaking and inversion, ST-segment elevation, and pathological Q-waves [11]. Figure 1.7 and Figure 1.8 show the association between MI progression and pathological ECG manifestations in specific leads (lateral leads I and V6; inferior lead II) for a patient with an LCx coronary artery occlusion. Specifically, Figure 1.7 (a) shows the ECG signals before occlusion. With the onset of LCx artery occlusion, the lateral ECG leads show

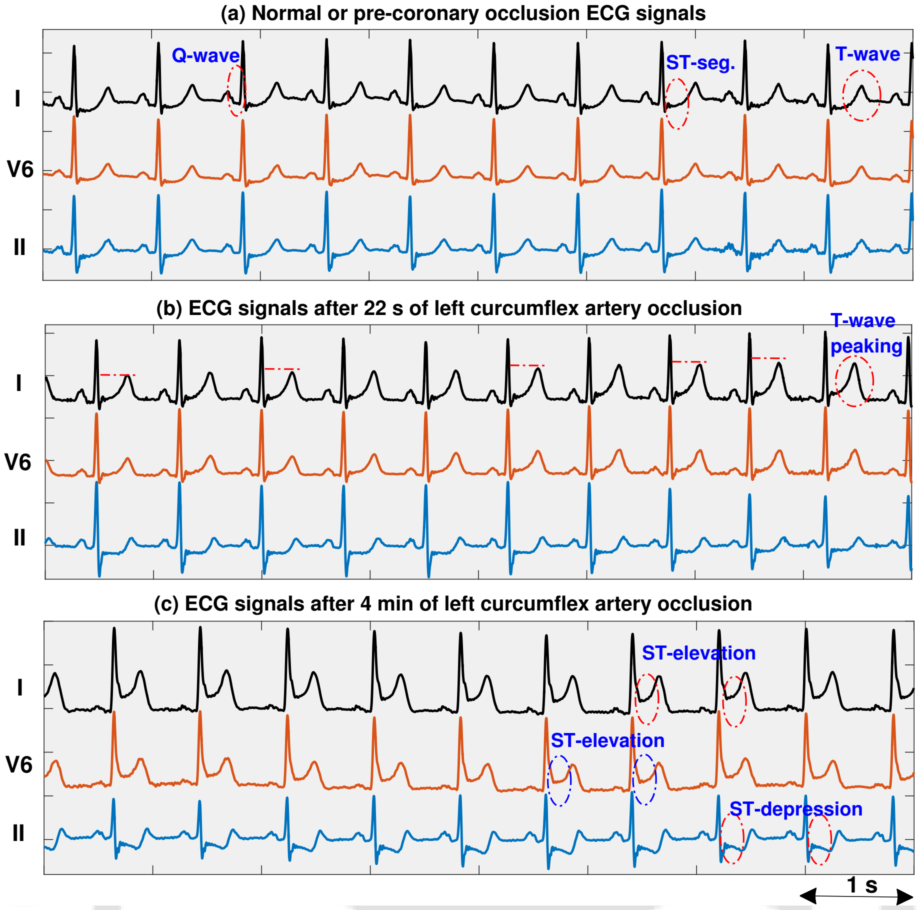


Figure 1.7: Typical ECG signals of a subject before and after left circumflex (LCx) coronary artery occlusion. The figure illustrates the dynamic changes in ECG characteristics during the disease progression at two lateral leads (I and V6) monitoring the infarcted region and one inferior lead (II) opposite the infarcted region. (a) Pre-coronary occlusion signals with normal ECG characteristics. ECG signals with pathological manifestations after 22 s of left circumflex coronary artery occlusion (b) and after 4 min of occlusion (c).

a gradual increase in the T-wave amplitude, i.e., T-wave peaking (Figure 1.7 (b)) with slight ST-segment elevations (Figure 1.7 (c)). Slight reciprocal ST-depressions appear in the inferior ECG lead II (Figure 1.7 (c)). Thus, T-wave abnormalities with slight ST-segment elevations are the pathological ECG features that help identify the early MI stage [24]. The same lateral ECG leads (I and V6) show the appearance of pathological Q-waves (due to the beginning of muscle necrosis), severe ST-segment elevations, and post-ischemic T-wave inversions during the acute MI stage (Figure 1.8 (a)). Finally, the appearance of deep pathological Q-waves with ST-segment recovery and T-wave inversions (Figure 1.8 (b)) helps identify the chronic MI stage. The careful analysis of these dynamic and subtle pathological ECG manifestations across the 12 leads is crucial for designing reliable automated MI severity classification approaches that can help prevent further disease progression by timely initiating suitable treatment plans [23, 25].

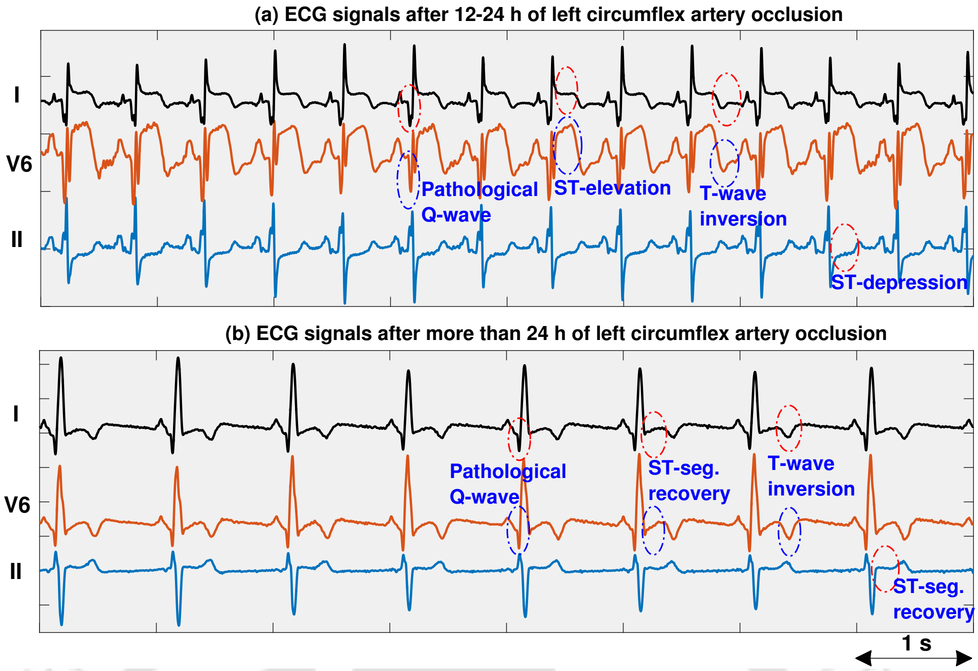


Figure 1.8: Illustrates the dynamic changes in pathological ECG characteristics during the disease progression. (a) After 12-24 h of left circumflex coronary artery occlusion. (b) After more than 24 h of occlusion.

1.2.2 Bundle Branch Blocks

Bundle branch blocks (BBBs) are the cardiac conduction system disorders where the electrical impulses from the AV node are blocked at left or right bundle branches [5, 17]. As a result, electrical signals travel more slowly on one side of the heart than on the other, affecting heart-pumping function. Often, BBBs are not life-threatening [26]. However, if left untreated, they can lead to severe complications such as sudden cardiac death, increased risk for heart failure, and fainting in the long run [26].

The BBBs show noticeable changes in the 12-lead ECG [10]. Specifically, right side chest leads V1/V2 and left lateral leads I/aVL/V5/V6 better characterize the disease [11]. During left BBB, depolarization of the left ventricle happens through the impulses spreading from the right ventricle. However, these electrical impulses must travel from myocyte to myocyte, a much slower process than traveling through the normal low-resistance conduction pathways. This slower or abnormal depolarization of the left ventricle results in a prolonged QRS duration (greater than 120 ms), deep and broad S-wave in lead V2, and broad notched R-wave in V6 (see Figure 1.9) [11]. In addition, ECG lead V6 shows ST-depression with T-wave inversions (Figure 1.9). Similarly, during right BBB, the abnormal depolarization of the right ventricle results in a prolonged QRS duration (greater than 120 ms), broad M-shaped rSR' QRS-pattern with positive T-waves in lead V2, and broad S-waves (greater than R-wave duration) with positive T-waves in lead V6 (see Figure 1.9) [11]. The early diagnosis of BBBs using automated ECG interpretation methods can help improve the

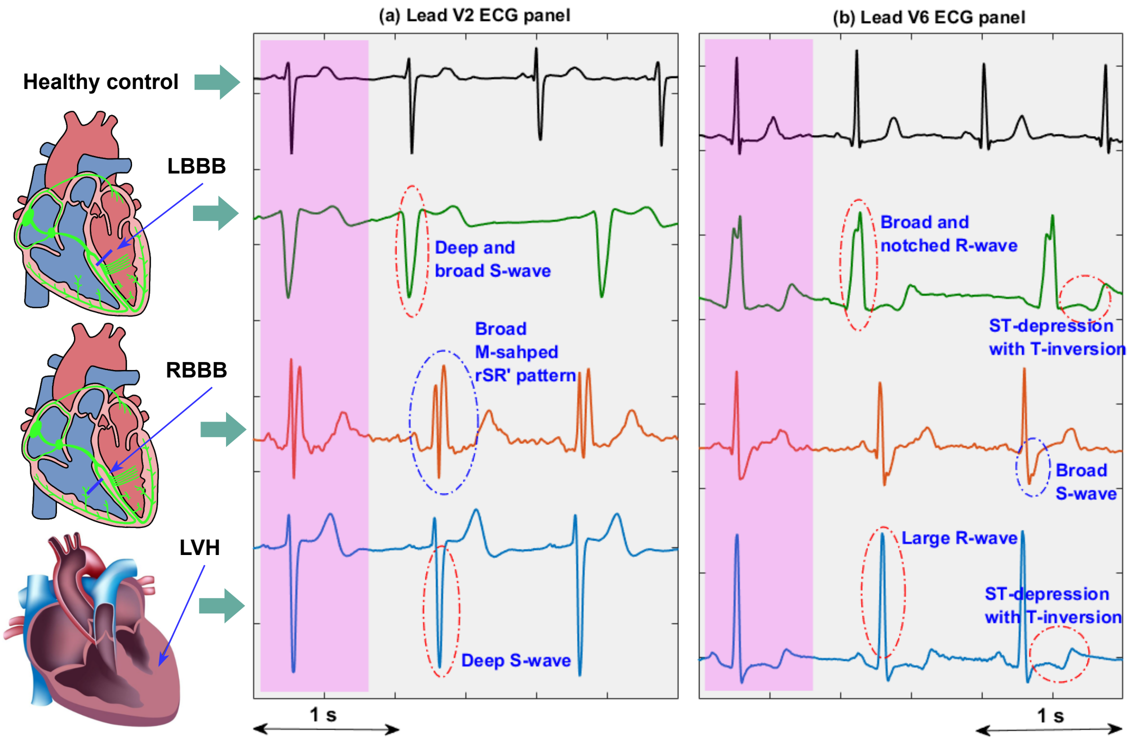


Figure 1.9: Comparison of (a) lead V2 and (b) lead V6 ECG signals for a healthy control subject and a left bundle branch block (LBBB), right bundle branch block (RBBB), and left ventricular hypertrophy (LVH) patients. The figure highlights the pathological ECG manifestations corresponding to LBBB, RBBB, and LVH patients.

coordination of ventricles by prompt initiation of cardiac resynchronization therapy (CRT) [27].

1.2.3 Hypertrophic Cardiomyopathy

Hypertrophic cardiomyopathy is a heart muscle disorder associated with abnormal thickening of the heart muscle, and it is more common in the left ventricle muscle, known as left ventricle hypertrophy (LVH) [5,28]. The thickened muscle becomes stiff, which reduces the amount of blood taken in and pumped out to the left ventricle. If left untreated, LVH can lead to severe complications such as mitral valve dysfunction, increased risk for heart failure, sudden cardiac death, and fainting [28]. Unfortunately, hypertrophy is the common reason for sudden cardiac death in young people and athletes under 30 years [29].

In clinical practice, 12-lead ECG is widely used to screen LVH patients [10]. During LVH, there is a larger heart muscle for the electrical impulses to pass through, leading to prolonged depolarization and delayed repolarization, which predominantly appear in the right-sided chest leads V1/V2 and left lateral leads I/aVL/V5/V6 [11]. Figure 1.9 shows characteristic ECG changes in leads V2 and V6 for the LVH patient. As can be seen, the increased depolarization voltages in the left ventricle manifest as slightly prolonged large R-waves in lead V6 and deep S-waves in lead V2. In addition, the delayed repolarization results in

ST-segment depression with T-wave inversions in lead V6. The effective handling of these pathological ECG variabilities using automated methods can assist clinicians in initiating suitable procedures that could increase the survival rate in hypertrophic patients [28].

1.2.4 Atrial Fibrillation

Atrial fibrillation (AF) is the most common irregular heart rhythm, which occurs due to uncoordinated atrial contractions [17]. During AF, instead of the SA node directing the electrical conduction system, many abnormal impulses rapidly fire the atria, causing them to depolarize irregularly. Because of this, the atria quiver instead of beating effectively to squeeze blood into the ventricles. The rapid impulses from the atria also reach the AV node, causing irregular ventricular contractions. AF is a rhythm disorder that can progressively worsen over time [30]. In the early stages, AF can be intermittent or paroxysmal and asymptomatic. If left untreated, the rapid and irregular AF rhythm can severely affect the blood flow dynamics and increase the risk of blood clots, stroke, heart failure, and cognitive impairments [30]. Depending on the severity of the AF condition, stroke risk can be substantially reduced by timely initiating appropriate treatment strategies such as oral anticoagulation, catheter ablation, and electrical cardioversion [31].

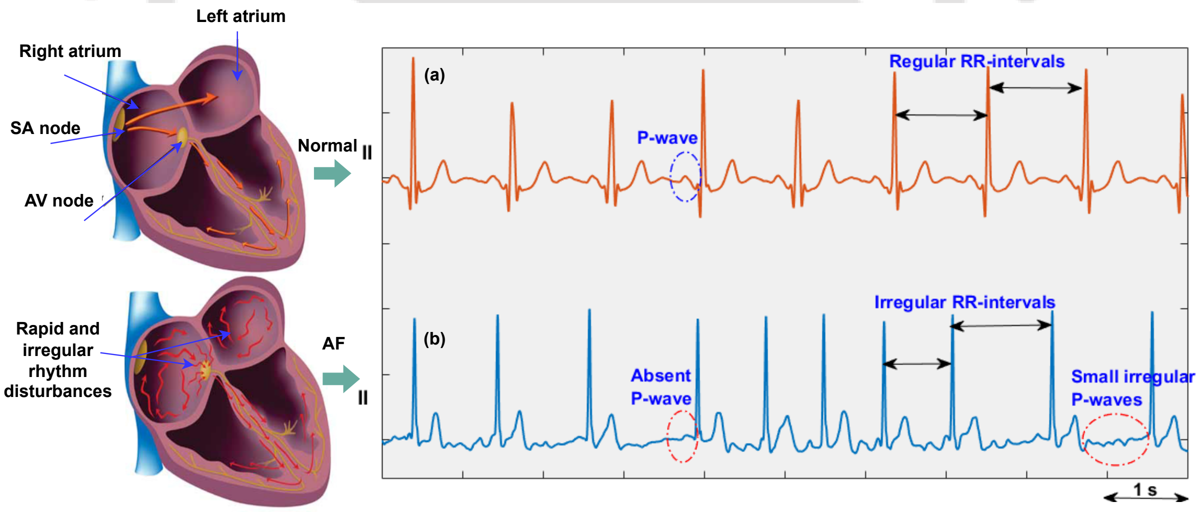


Figure 1.10: Comparison of normal sinus rhythm and atrial fibrillation ECG signals. (a) Healthy control signal with normal ECG characteristics. (b) Atrial fibrillation ECG signal with abnormal ECG patterns such as the absence of P-waves, presence of f-waves, and irregular RR-intervals.

The single-lead ECG has been an indispensable tool for the early diagnosis and management of AF rhythm disorder [31]. On the ECG, the absence of P-waves and the presence of fibrillatory waves (f-waves), and an irregular ventricular rhythm or RR-intervals characterize the AF condition (see Figure 1.10 (b)). The recent developments in portable and wearable ECG devices have led to early and effective AF screening via remote monitoring; thus, automated methods suitable for these devices require rigorous investigation [32].

1.2.5 Congestive Heart Failure

Congestive heart failure (CHF) is a complex clinical syndrome in which the heart fails to pump enough oxygenated blood to meet the body's circulatory demands [33]. It can result from any cardiac disorder that impairs ventricular filling or pumping of the blood, including MI, HYP, CD or BBBs, valvular heart disease, and cardiac arrhythmia [33]. During CHF, the blood often backs up, and fluid can build up in the lungs, feet, ankles, and legs, causing shortness of breath [34]. CHF can be life-threatening; thus, proper treatment procedures addressing the underlying heart disorders or diuretics, angiotensin converting enzyme (ACE) inhibitors, and vasodilators help improve the signs and symptoms of heart failure [34].

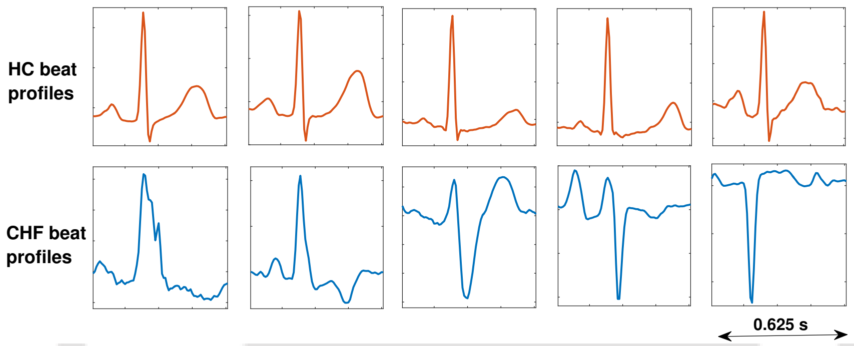


Figure 1.11: Typical ECG beat profiles of healthy control (HC) and congestive heart failure (CHF) subjects.

The single-lead ECG is the primary non-invasive test often preferred by clinicians to diagnose CHF [35]. The impairment of heart function during CHF causes the reduced heart rate variability (HRV), derived from the RR-intervals (duration between successive R-waves) of the ECG [35]. Because the etiology of the CHF is diverse, there are no definite ECG clinical guidelines for its diagnosis apart from RR-interval variability [34]. However, specific ECG patterns, including ST-segment elevation or depression, prolonged QRS- and QT-duration, low-amplitude QRS-complex, deep S-wave, and T-wave abnormalities, are often associated with the CHF [35, 36], as shown in Figure 1.11. The automated analysis of these complex pathological ECG manifestations can help screen CHF patients efficiently.

1.2.6 Multimorbidity

Multimorbidity is the co-occurrence of two or more cardiac disorders such as MI, HYP, BBBs, and AF in the same individual [37]. The presence of multiple cardiac disorders simultaneously in an individual has a strong association with in-hospital mortality and other adverse outcomes [38]. Depending on the type and number of cardiac disorders present at the same, timely initiation of multi-modal treatment procedures may help improve clinical outcomes [38].

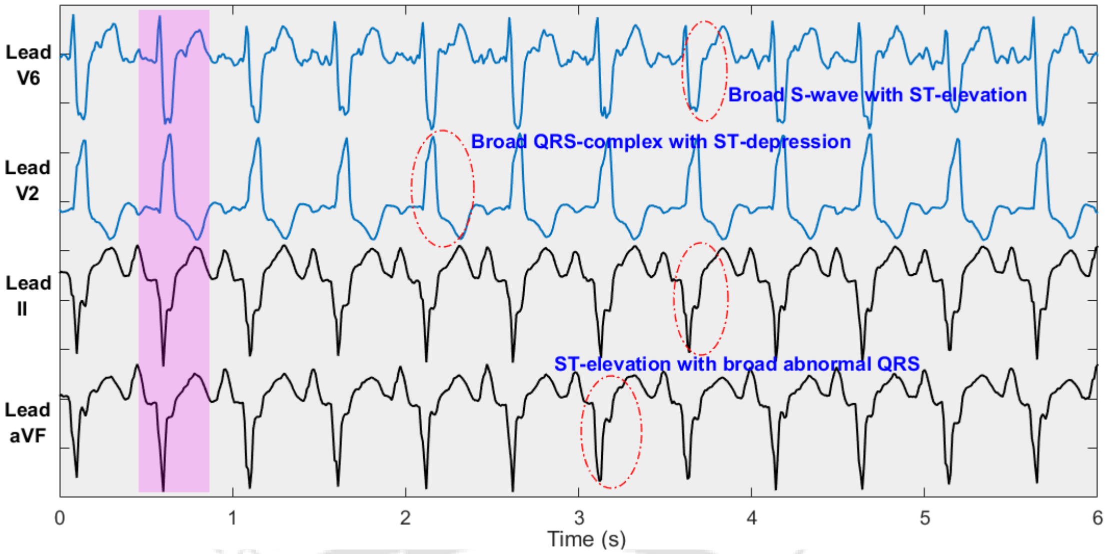


Figure 1.12: Illustrates the abnormal ECG manifestations in the ECG leads V6, V2, II, and aVF for a multimorbidity patient with inferior-lateral MI and right BBB co-occurring diseases. The figure highlights the complex interaction between the pathological ECG morphologies of two cardiac disorders.

Apart from helping analyze single cardiac disorders, the standard 12-lead ECG also shows visible morphological changes for patients with multiple co-occurring cardiac disorders [39]. Figure 1.12 depicts the abnormal ECG changes in the leads V6, V2, II, and aVF for a multimorbidity patient with inferior-lateral MI and right BBB. As can be seen, leads V6 and V2 show the right BBB ECG changes and MI-related ST-elevations (in lateral lead V6). Similarly, inferior leads II and aVF show the ST-segment elevations related to MI and abnormal and broad QRS changes related to right BBB. Although 12-leads show complex pathological ECG changes, its careful analysis may help in the early detection of multimorbidity patients.

In the following paragraphs, we present some challenges that affect the effectiveness of automated methods for diagnosing cardiac abnormalities from the 12-lead or single-lead ECG signals.

Variabilities in the pathological ECG manifestations of the cardiac disorders: Because of the diverse etiology of the cardiac disorders (coronary artery: MI, electrical conduction system: BBBs, cardiac muscle: LVH, and rhythm: AF and CHF), the pathological ECG manifestations widely vary in terms of shape, duration, amplitude, and temporal association [11,39]. Moreover, based on the severity of disease progression, these manifestations vary dynamically and undergo subtle changes within the same cardiac abnormality [10]. The effective handling of these subtle variabilities is often challenging for developing reliable automated ECG analysis methods for the diagnosis and severity assessment of cardiac disorders [13].

Diagnostic redundancy of 12-lead ECG: As a simultaneous measurement of cardiac electrical activity from different spatial angles, the 12-lead ECG can diagnose various localized cardiac ailments. However, depending on the type of cardiac ailment, distinct pathological ECG manifestations appear in specific leads of the 12-leads, with others being clinically redundant [13]. This implicit redundancy of the 12-lead ECG often hinders the clinically relevant information, affecting the efficacy of automated diagnosis systems.

Intermittent nature of the disease: The long-term ECG recordings aid in the early diagnosis and characterization of paroxysmal AF [9]. However, the pathological manifestations of paroxysmal AF vary dynamically and occur sporadically on the long-term ECG recording with more than 90% of normal rhythm [9, 14]. It is, therefore, difficult to interpret these under-representative, short, and dynamic ECG patterns of paroxysmal AF from the long-term recordings using automated methods.

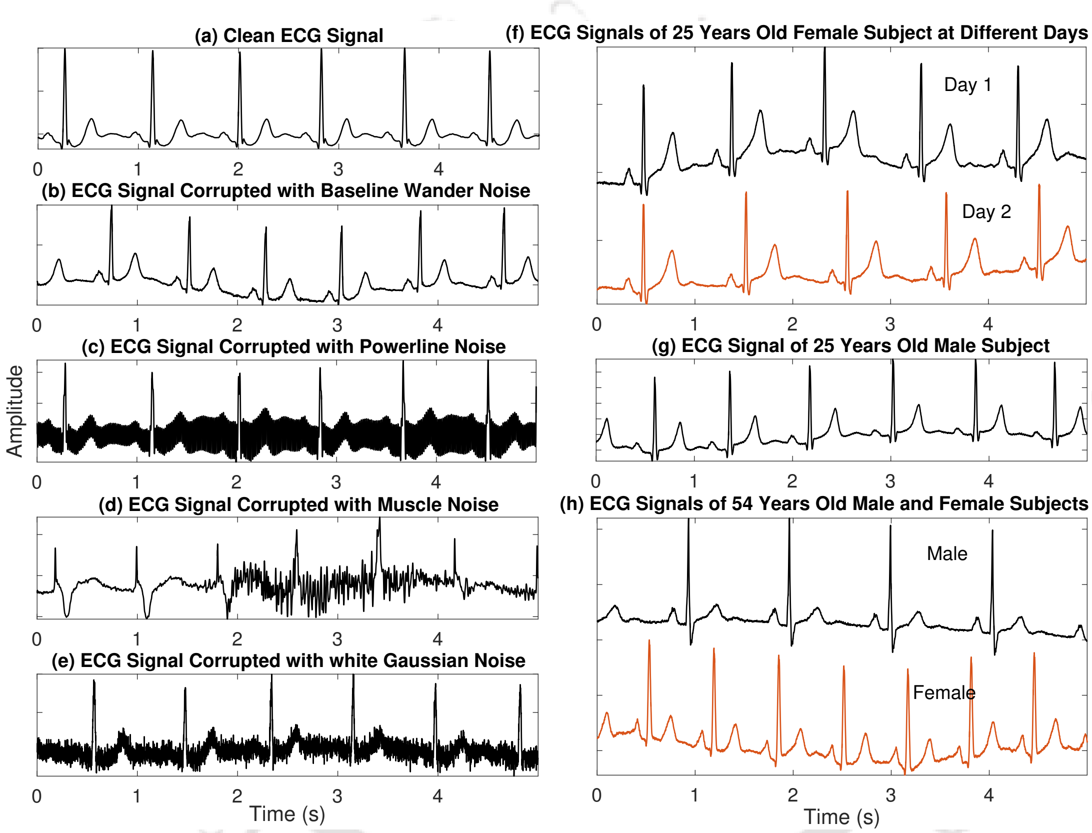


Figure 1.13: Different types of noises obscure the ECG morphological characteristics. (a) Clean ECG signal, (b) ECG corrupted with respiratory baseline wander noise, (c) ECG corrupted with powerline interference, (d) ECG corrupted with muscle contraction noise, and (e) ECG corrupted with white Gaussian noise. The figure also illustrates the intra- and inter-subject variabilities of ECG signals. (f) ECG signals of 25 years female subject on different days, (g) ECG signals of 25 years male subject, and (h) ECG signals of 54 years old male and female subjects.

Effect of noises on ECG signal: In practice, the relatively low amplitude ECG signals are often corrupted with various artifacts, including low-frequency baseline wander (BW) due to patient movement and respiration noise, power line interference noise, and high-frequency muscle contraction noise [8]. The ECG signals contaminated with some of these noise sources are shown in Figure 1.13. As can be seen, the presence of the noises obscures the morphological characteristics of the ECG, which pose a significant challenge for the automated diagnosis methods to provide robust clinical decisions [40]. In addition, the inter-subject (age, gender, heart rate variability, and body mass index) and intra-subject (physiological and mental status) ECG variabilities further adds to the difficulty of designing robust automated methods [41].

The following section presents different approaches used in literature for the automated diagnosis of cardiac disorders from single- and 12-lead ECG signals.

1.3 Automated Diagnosis of Cardiac Disorders from ECG - A Review

The overarching aim of ECG-based automated diagnosis systems is to extract, characterize, and recognize the diagnostic information about cardiac disorders from ECG signals [8, 12]. Such systems can facilitate rapid and improved objective clinical decision-making, minimize medical errors, and optimize clinical workflow [13]. A general block diagram of the ECG-based automated diagnosis system is depicted in Figure 1.14. As can be seen, there are two main stages of operation, i.e., the training stage and the prediction or testing stage. At the training stage, the labeled ECG data are acquired, preprocessed, feature engineered, and the classification model is learned. At the testing stage, the ECG signals from unseen patients are preprocessed, feature engineered and fed to the trained classifier model to output the diagnosis decision. Both stages include the same preprocessing and feature engineering steps.

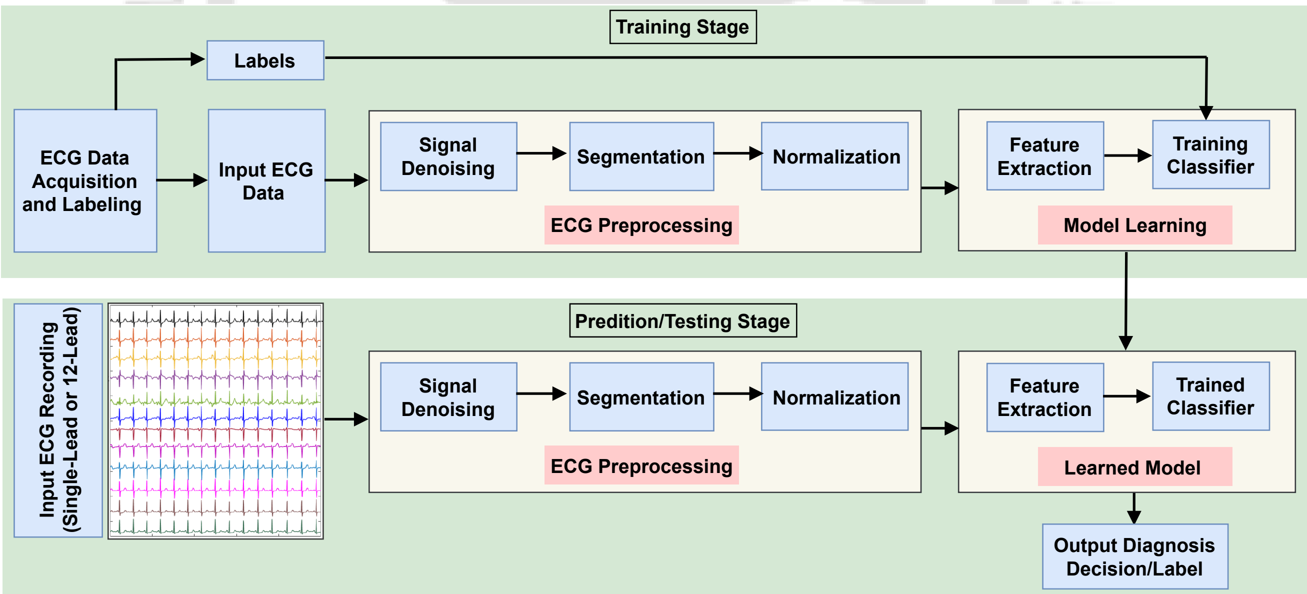


Figure 1.14: General block diagram of an ECG based automated diagnosis system with training and testing stages.

The training phase starts with acquiring clinical ECG data from a large, retrospective cohort of patients. Depending on the clinical application, data acquisition includes either single- or 12-lead ECG signal extraction. The acquired dataset then goes through the most complex and crucial annotation or labeling process. Here, each ECG recording is annotated by the consensus of a committee of board-certified, actively practicing cardiologists as either normal or abnormal types based on the observed changes in the morphological ECG characteristics [42].

The preprocessing of the ECG includes signal denoising, segmentation, and normalization [8]. The following paragraphs present the details of these steps.

Signal denoising: The presence of various noise artifacts, including respiratory noise, muscle contraction noise, motion artifacts, and powerline interference, obscures the morphological characteristics of the ECG, which often leads to the wrong clinical interpretation of the signal [8, 40]. Hence, the ECG denoising is considered in the preprocessing stage to remove noises and enhance signal quality before performing any disease analysis. The most common denoising techniques explored in the literature include band-pass filters, notch filters, adaptive filters [8], wavelet transform [43, 44], empirical wavelet transform (EWT) [45], empirical mode decomposition (EMD) [44], and variational mode decomposition (VMD) [46].

Segmentation: ECG signals undergo segmentation after denoising, which refers to the analysis window of input used for the diagnosis. The segmentation techniques used in the automated diagnosis systems can be categorized into two classes: (i) beat-based and (ii) rhythm-based. Beat-based segmentation requires R-peak detection followed by a fixed window around the R-peak to extract P-QRS-T waves, representing a heartbeat having intra-beat information. Such segmentation has been explored in the literature to diagnose ectopic beats [40], MI [47, 48], and CHF beats [49]. The frame-based segmentation selects a blind excerpt from the ECG recording for the analysis, thereby alleviating the need for the R-peak detection. The segmentation of single-lead ECG consists of intra-lead rhythm information, whereas segmentation of 12-lead ECG consists of both intra-lead and inter-lead diagnostic information. It has been a standard technique widely investigated in the existing studies to diagnose various cardiac ailments [50–53].

Normalization: The last step of the prepossessing stage is normalizing the segmented ECG signals. In practice, the amplitudes of the ECG recordings vary among different subjects and acquisition devices. Therefore, the min-max and Z-score normalization techniques (see Appendix A) are commonly used for amplitude scaling that standardizes ECG signals for analysis. These techniques normalize the ECG amplitudes into a range of 0 and 1 or to have zero mean and unit variance, respectively [54].

The preprocessed ECG signals and corresponding labels are next fed to the model learning block, which learns the required parameters in a supervised manner to improve the diagnosis performance. Over the past few decades, researchers have developed different model learning approaches for ECG-based automated diagnosis. Early works have employed machine learning (ML) based approaches consisting of feature extraction followed by classification [47, 48, 50, 51, 55]. The feature extraction is performed under the guidance of medical experts, where various clinically informative features that comprehensively characterize the given cardiac abnormalities are extracted from the preprocessed ECG. The ML classifiers take these features as input and provide the output diagnosis label. The ML-based approaches heavily rely on domain expert knowledge to engineer valuable features from the ECG signals, which are subjective and face the challenge of feature quality and robustness [54]. In recent years, deep learning (DL) based approaches have been increasingly used for ECG-based automated diagnosis [15, 54, 56]. The DL models, popularly known as representation learning models, consist of multiple processing layers. Each layer can learn increasingly abstract, high-level feature representations or patterns automatically form the ECG signals for their effective classification, thereby eliminating the burden of handcrafted feature engineering. The models demonstrate improved diagnosis and better generalization performance than the ML methods [57]. This dissertation focuses on automated analysis of various cardiac disorders, including MI, BBBs, HYP, CHF, and AF. The following sections briefly describe the existing ML- and DL-based approaches for diagnosing these cardiac disorders from the single/12-lead ECG signals.

1.3.1 Existing Myocardial Infarction Diagnosis Methods

The MI can be characterized by the appearance of T-wave peaking and inversion, ST-segment elevation, and pathological Q-waves in two or more anatomically contiguous leads of the 12-lead ECG [11]. Over the last few decades, researchers have proposed various ML- and DL-based approaches for objectively quantifying these pathological ECG changes for the diagnosis.

ML-based approaches: The ML-based methods mainly rely on extracting clinically informative features from the pathological manifestations of the ECG signals, followed by classification using conventional ML classifiers. Early studies adopted expert morphological features associated with the ECG waveform for the diagnosis [55]. For instance, Arif *et al.* [50] computed several time-domain features such as Q-wave amplitude and ST-segment deviation from the 12-lead ECG to train a K-nearest neighbor (KNN) classifier for the MI detection. Similarly, the authors of [58] evaluated the polynomial fitting features of the ST-segment followed by multiple instance learning (MIL) classifier for identifying MI patients from the healthy control (HC) subjects. It is worth emphasizing that the performance of such morphological features highly depends on the accuracy of ECG fiducial points (P, Q, R, S, and T) detection algorithms, where only R-peak detection algorithms are reliable enough to be used in the real-world applications. Thus, methods based on such features often generate false MI alarms in the presence of noises [59]. In subsequent studies, instead of explicitly quantifying pathological changes from the ECG, researchers have explored the use of signal transformation techniques, such as discrete Fourier transform (DFT) [47], wavelet transform (WT) [48, 51, 60–62], higher order singular value decomposition (HOSVD) [63], and empirical wavelet transform (EWT) [64] that can implicitly represent the diagnostic information indicative of MI and allow effective feature extraction. The extracted features are then fed to the decision tree [47,48], KNN [50,51,60], logistic regression [47], support vector machines (SVM) [51,58,60,61,63,64], random forest [58,65], and artificial neural network (ANN) [66] for classifying MI and HC ECG recordings. Although these methods show promising performance on smaller datasets, they have two crucial limitations. (i) The design of handcrafted features is highly subjective to the designer's domain expertise, and the extracted features are susceptible to the various artifacts and intra- and inter-subject ECG variabilities [41,67]. (ii) Considering the progressive nature of MI and dynamic changes in the pathological ECG manifestations, it is often difficult to find a fixed set of informative and discriminative features to train the ML classifiers [10].

Recently, DL-based approaches have shown impressive recognition results in ECG signal analysis [15,54,57]. These approaches effectively combine feature extraction and classification through end-to-end learning, thereby alleviating manual feature extraction. Specifically, the DL models have shown promising MI detection results over ML methods by effectively learning the variations in the 12-lead ECG. The following section presents related literature on the DL-based methods for MI diagnosis.

DL-based approaches: One of the most common DL models that are being increasingly used for classifying ECG signals is convolutional neural networks (CNNs) [68]. CNNs have proven effective in recognizing crucial ECG patterns, such as P-waves, QRS-complexes, ST-segment, and T-waves, thereby improving classification performance [69]. A typical CNN architecture consists of three primary layers: convolutional, pooling, and fully-connected (FC) layers [68]. The initial stages of the CNNs use multiple convolution layers, non-linear activations, and pooling layers to systematically process the input ECG signals and automatically learn valuable features for the classification end-to-end. The convolution layer is the core building block of CNNs, which consists of multiple convolutional kernels that convolve with the input ECG and extract various characteristic representations of the raw input known as feature maps [68,70]. The kernels are nothing but a group of 1D filters, which are essentially trainable weights. These learnable convolutional filters effectively encode the local morphological features of the input ECG signals [68]. Activation functions generally follow the output feature maps of the convolution layers to add non-linearity to the features. The pooling layers are then adopted to merge systematically similar features from the local patches of feature maps to a more robust one. Thus, pooling layers output robust feature maps invariant to the translation by condensing them to a smaller size. By convention, multiple stacks of convolution, activation, and pooling layers are used to extract powerful features. Finally, these features are transformed into a 1D vector and fed to the output FC layers for classification. All the weights of the CNNs are trained iteratively using an error backpropagation [68]. For more details on different CNN layers, see Appendix A.

The first use of 1D-CNNs for ECG signal analysis was reported in 2016 by Kiranyaz *et al.* [68]. The authors have employed CNNs for patient-specific arrhythmia classification from ECG signals. However, such a patient-specific model cannot be applied to unknown patients, limiting the method's applicability in real-world applications. Specifically, some studies recently proposed CNN-based models for MI diagnosis. Acharya *et al.* [69] proposed an 11-layer deep CNN (DCNN) architecture for MI detection using single-lead (lead II) ECG beats. A similar DCNN model was also explored in [71] for MI diagnosis from lead V4 ECG beats. To utilize more diagnostic information, the authors of [72] employed a segment of three ECG leads (II, III, and aVF) for identifying inferior MI (IMI) patients. The authors developed an inception-based CNN model for IMI diagnosis and achieved an accuracy of 84.54%. Liu *et al.* [73] designed a multilead-CNN (ML-CNN) architecture using four ECG leads (V2, V3, V5, and aVL) for classifying anterior MI and HC ECG recordings. The method uses a novel sub 2D-convolution and Lead Asymmetric Pooling (LAP) techniques to optimize CNNs, which resulted in an improved classification accuracy of 96%. With the same ECG-leads combination, authors in [74] developed a residual DCNN model, slightly improving anterior MI detection accuracy over [73]. Similarly, a multi-channel lightweight CNN model was proposed in [75] to improve the detection performance of anterior MI using three ECG leads (V1, V2, and V3).

The methods discussed above utilize a subset of 12-lead ECGs for diagnosing localized MIs such as interior/anterior MI. Such apriori selection of a set of ECG leads for specific MIs, limits these method's applicability in practical diagnosis systems. Because in real-world, patients can present with MI occurring at any spatial location (inferior, anterior, lateral, and septal) of the heart. To improve the reliability of MI diagnosis, Strodthoff *et al.* [76] employed the standard 12-lead ECGs as an input to the residual deep CNN model. The use of comprehensive diagnostic information from twelve different views (12-lead ECG) improved MI detection accuracy. However, the residual deep CNN model processes the input 12-lead ECG as a 2D-data (timestamps, ECG leads), thus not fully exploiting the crucial lead-specific information of the 12-lead ECG. To mitigate this issue, Liu *et al.* [77] presented a multiple-feature-branch CNN (MFB-CNN) architecture that comprehensively represents the 12-lead ECGs for the diagnosis. The MFB-CNN consists of multiple feature branches to extract features from each ECG-lead and a global fully connected Softmax layer to summarize all the feature branches, determining the final diagnosis decisions. The method achieved an improved MI detection accuracy of 99.9%. The same research group [78] further proposed a hybrid model combining CNNs and bi-directional recurrent neural networks (BRNN) called MFB-CBRNN for classifying MI and HC 12-lead ECG recordings. The method utilized an MFB-CNN structure for feature extraction and a BRNN for feature summarization, followed by a Lead Random Mask (LRM) optimization technique to achieve robust classification performance.

1.3.2 Existing Congestive Heart Failure Diagnosis Methods

The abnormal heart function during CHF causes autonomic nervous system (ANS) dysfunction that can be characterized by the variations in the heart rate variability (HRV) derived from the RR-intervals [79]. Moreover, considering the diverse etiology of the CHF, specific ECG patterns, including ST-segment elevation or depression, prolonged QRS- and QT-duration, low-amplitude QRS-complex, deep S-wave, and T-wave abnormalities, are often associated with the CHF [35, 36]. Several researchers have explored these abnormal ECG changes for detecting CHF patients using ML- and DL-based approaches.

ML-based approaches: Existing CHF detection methods mainly employed ML-based frameworks consisting of feature extraction and classification [80]. In the feature extraction stage, most studies focused on using either standard long-term HRV [81, 82] or short-term HRV [83–85] indices to characterize the CHF patients. The HRV indices can be broadly divided into time-domain and frequency-domain. The popularly explored time-domain indices include the average of all normal RR-intervals (AVNN), the standard deviation of all normal RR-intervals (SDNN), the root-mean-square of successive RR interval differences (RMSSD), and the percentage of successive RR intervals that differ by more than 50 ms (pNN50). The frequency-domain indices were extracted from the power spectral density of RR-interval data, which include total spectral power up to 0.4 Hz (TOTAL POWER), total spectral power between 0.04-0.15 Hz (LF), total spectral between 0.15-0.4 Hz (HF) and the ratio of LF-to-HF power (LF/HF). Generally, HRV feature-based approaches require long-term ECG signals of more than 24 h duration or at least 5-min/30-min duration for the reliable assessment of CHF disease. This requirement limits these method's applicability in the remote diagnosis systems using portable ECG devices with storage constraints [86]. In addition, HRV features were only effective in identifying severe CHF patients (New York Heart Association (NYHA) III and IV) with acute ANS damage and thus are not reliable for diagnosing patients at the early stages of CHF (NYHA I and II) [79]. To mitigate these issues, instead of using long-term RR-interval data, recently, few studies have employed raw ECG segments of shorter duration for the CHF diagnosis. In particular, signal transformation techniques like wavelet transform [52, 53] and autoregressive model [87] have been applied to the ECG signals to grossly extract features related to the pathological ECG manifestations of CHF condition. The extracted features were provided to the conventional ML classifiers, including decision trees [52, 81–84], KNN [82], RF [85, 87], and sparse-representation [53] for classifying CHF and HC subjects. Because CHF is a complex clinical syndrome, its diverse pathological ECG changes may not be quantified well with the above-discussed feature engineering-based methods.

DL-based approaches: Acharya *et al.* [88] and Porumb *et al.* [49] have developed deep CNN-based architectures for discriminating CHF patients from HC subjects. Both approaches used a stack of CNN layers to learn the complex pathological ECG manifestations for reliable CHF diagnosis. Similarly, the hybrid model using deep CNNs and RNNs was proposed in [89] for CHF analysis. The method utilized CNNs for feature extraction followed by RNNs for feature summarization to improve CHF vs. HC classification using ECG signals. It is worth mentioning that, along with the ECG classification, Porumb *et al.* [49] have also presented preliminary results on the interpretation of the designed CNN-based network using the Gradient-weighted class activation mapping (Grad-CAM). The Grad-CAM helps visualize the salient ECG features contributing most to the network's assignment of the labels (CHF or HC) at the testing stage.

1.3.3 Existing Atrial Fibrillation Diagnosis Methods

On the ECG, AF can be characterized by irregular RR-intervals and the absence of P-waves or the presence of f-waves. The currently accepted convention for AF diagnosis is the presence of an episode lasting at least 30 s on the ECG; thus, its diagnosis is categorical, meaning that eventually, a patient is classified as AF or non-AF [31]. Over the years, several ML- and DL-based approaches have been investigated to classify AF and non-AF events from the single-lead ECG signals.

ML-based approaches: ML-based AF detectors can be broadly divided into rhythm-based, rhythm- and morphology-based. Rhythm-based detectors only exploit the increased RR-interval variability associated with AF rhythm [90]. On the other hand, rhythm- and morphology-based detectors exploit both the RR-interval and P-wave morphological (absence of P-waves or presence of f-waves) information of the ECG to diagnose AF [90]. In the following paragraphs, we provide an overview of these approaches. *Rhythm-based detectors:* Because R-peak fiducial points can be robustly detected from the ECG signal, most existing methods have focused on the analysis of RR-intervals for detecting AF. Early works have employed the Markov process [91] and cumulative distribution [92] of the RR-intervals to characterize the irregular rhythm. In subsequent studies, the 2D-representation of RR-intervals, including the Poincare plot [93] and Lorenz plot [94, 95], were explored to improve the AF detection performance. In addition, the authors in [96, 97] have investigated the RMSSD features, which have shown to be good at classifying AF and non-AF patients. With the advancements in information theory, the use of complexity measures like Shannon entropy [96–98] and sample entropy [99] led to the improved quantification of the dynamic RR-interval changes of AF. Few studies even used the combination of multiple non-linear features with SVM [100] and RF [101, 102] to improve the AF detection performance. The above discussed rhythm-based methods increase false AF alarms in the presence of ectopic beats (premature atrial and ventricular contractions (PACs/PVCs)) [102]. Also, the performance highly depends on the robustness of the R-peak detection algorithms.

Rhythm- and morphology-based detectors: In order to improve AF detection in the presence of ectopic-beat rhythms, a combination of both RR-interval and P-wave morphological information has been explored in the literature. For instance, in [103], RR-irregularity, f-waves presence, P-waves absence, and noise-levels were combined with a fuzzy logic classifier to detect paroxysmal AF. To analyze the f-waves information, a few researchers first extracted the atrial activity signal by canceling QRST on the ECG signal [104], followed by spectral analysis of the atrial signal to extract features related to the absence of P-waves or the presence of f-waves [105–108]. In particular, atrial activity features and various RR-interval-based statistical and morphological features were combined with ML classifiers for efficiently classifying AF and non-AF rhythms [109, 110]. Few studies explored wavelet [111] and variational mode decomposition [112] techniques to improve AF diagnosis by grossly extracting robust features. It is worth emphasizing that features from the low-amplitude atrial signals were commonly obscured by noise, leading to similar or degraded performances as that of rhythm-based detectors [90].

DL-based approaches: Early DL works [113, 114] have utilized RR-interval data combined with DCNN models to learn irregular rhythm information associated with the AF. The authors in [115] transformed the RR-interval data into a 2D Poincare plot and employed the VGGNet model to improve the AF prediction. To alleviate the need for signal transformation techniques, the authors in [90, 116, 117] have investigated a 1D-CNN model with raw ECG data for classifying AF from non-AF rhythm. In order to improve the AF detection in the presence of other rhythms, recently Fan *et al.* [118] have employed ECG data combined with a two-stream multi-scale CNN (MS-CNN) model. The MS-CNN operates at different kernel sizes, effectively extracting and combining various multi-scale ECG features for improving AF diagnosis. The method generates state-of-the-art AF diagnosis performance with an accuracy of 98.1$\%$.

Till now, we have presented the existing methods for disease-specific (MI/CHF/AF) analysis of ECG. However, in real-world clinical settings, patients can present with any cardiac disorder, including MI, CD, HYP, AF, and so on. Therefore, the automated diagnosis systems should be able to analyze ECG signals of multiple cardiac abnormalities for their diagnosis and appropriate management [13, 39]. Over the years, researchers have explored several methods for detecting multiple cardiac disorders, referred to as multi-class ECG classification approaches. Multi-class ECG classification means a classification task with more than two classes, assuming that each ECG record corresponds to only one cardiac disorder/label [119]. Moreover, some patients may present with co-occurrence of two or more cardiac disorders simultaneously, known as multimorbidity patients [38]. Identifying multimorbidity patients from the ECG is referred to as a multi-label ECG classification problem, and it assigns a set of target labels to each ECG record [119]. The following section presents an overview of the existing approaches for multiple cardiac ailments classification.

1.3.4 Existing Multiple Cardiac Disorders Classification Methods

This dissertation focuses on the simultaneous analysis of life-threatening cardiac disorders, such as MI, CD or BBBs, and HYP using ECG signals. In the following paragraphs, we present existing multi-class and multi-label ECG classification methods related to these disorders.

Multi-class ECG classification approaches: Most existing methods employed 12-lead ECG to classify localized cardiac disorders, such as MI, CD or BBBs, and HYP. The appearance of ST-elevation, T-wave peaking and inversion, and pathological Q-waves characterizes the MI disease. The broad and bizarre QRS-complexes, deep and broad S-waves with ST-depressions, and T-inversions in the right-sided and lateral ECG leads are often associated with the BBB disorder. Similarly, the increased R-wave amplitudes with ST-depression and T-inversions in the lateral ECG leads and deep S-waves in the right-sided ECG leads help diagnose HYP. It is worth emphasizing that BBB and HYP are commonly known as "mimicking MI" diseases owing to their MI-like pathological ECG features [120]. Thus, it is often difficult to simultaneously analyze these disorders, and only a few works have been explored in this direction. For instance, Tripathy *et al.* [121] proposed an ML-based approach consisting of multi-scale phase alternation features with a KNN classifier to classify HC, BBB, HYP, and MI heart conditions from the 12-lead ECG. A similar study combining morphological and temporal features with an RF classifier has been investigated in [122] for classifying HYP and non-HYP 12-lead ECG signals.

In the direction of DL-based models, Hannun *et al.* [15] developed a DCNN-based model to classify NSR and eleven irregular heart rhythms based on single-lead ECG. The method exceeded the performance of cardiologists in their experiments. A hybrid model combing DCNN and LSTM was proposed in [123] for classifying NSR and three irregular rhythms using 12-lead ECG. The above methods analyze only rhythm disorders, and their performance on the acute localized disorders required verification. Ko *et al.* [124] designed a DCNN-based model to classify HYP and non-HYP 12-lead ECG and achieved an area-under-curve (AUC) of 96%. In 2018, China Physiological Signal Challenge 2018 (CPSC2018) released a nine-class 12-lead ECG dataset of 6,877 recordings [125]. The nine classes include NSR, AF, first-degree AV block (I-AVB), left BBB, right BBB, PVC, PAC, ST-depression (STD), and ST-elevation (STE). The CPSC2018 dataset provided an impetus to the progress of multi-class ECG classification of acute cardiac disorders. Consequently, a series of hybrid DL-models combing CNN, LSTM or bidirectional gated recurrent units (BiGRU) and attention modules have been investigated in [126–128] for the multi-class ECG classification. The attention module in these methods automatically attends to the informative CNN+LSTM [127] or CNN+BiGRU [126, 128] feature maps for efficient classification. In order to learn the multi-scale pathological ECG features of cardiac ailments, Wang *et al.* [129] designed a two-stream deep multi-scale fusion network (DMSFNet). The DMSFNet model fuses the multi-scale feature outputs from two deep CNNs at the FC+Softmax layer for effective multi-class ECG classification.

Multi-label ECG classification approaches: Recently, few researchers have presented their preliminary studies on the multi-label ECG classification using a small dataset of 476 multi-label ECG recordings from the CPSC2018 dataset. For instance, Li *et al.* [130] presented an ML-based multi-objective optimization model that exploits the disease correlations to classify the multi-label ECG signals. The method used 117 handcrafted features followed by classification and achieved an average F1-score of 60.8%. This poor performance of ML methods can be improved using DL-based methods. For instance, Jia *et al.* [131] proposed a voting strategy-based ensemble CNN model for multi-label ECG classification. The method achieved an impressive F1-score of 87.2% on the CPSC2018 dataset. A similar study based on DCNN was also investigated in [132] to predict the presence of co-occurring cardiac disorders from the 12-lead ECG beat matrix. Finally, a hybrid model combing CNN, BiLSTM, and attention module has been developed in [133] for multi-label ECG analysis. However, the method employed only single-lead ECG for the classification, which resulted in poor performance with an average F1-score of 60.1%.

Further details of the existing ML- and DL-based approaches for diagnosing cardiac disorders are provided in Appendix A as comparison tables.

1.4 Motivation for the Research Work

Although several approaches have been investigated for the ECG-based automated diagnosis, a few research challenges associated with the analysis of ECG for reliable diagnosis and characterization of cardiac disorders have not been addressed and are given as follows.

- Even though MI is a progressive cardiac disorder, existing methods [55] consider it as one broad category for the diagnosis and do not address its severity staging. In practice, the severity assessment of MI in terms of early MI, acute MI, and chronic MI using ECG signals can play a crucial role in assisting cardiologists to timely initiate appropriate treatment strategies that can avert further disease progression and reduce associated mortality [21]. The subtle variabilities in the pathological ECG manifestations and dynamic changes in their temporal dependencies during the disease progression pose a challenge for reliable classification of MI severity stages. In addition, the implicit intra- and inter-lead diagnostic redundancies of the 12-lead ECG often obscure the clinically relevant information, which further adds to the difficulty of classifying severity stages. Therefore, we hypothesize that an automated method that can systematically process the 12-lead ECGs to represent better the temporal dynamics of subtle pathological ECG variabilities while alleviating/emphasizing the diagnostically redundant/relevant information could improve the performance of MI severity stages classification.

- As discussed before, the etiology of CHF is diverse. It results in wide variability in the pathological ECG features and dynamic changes in their temporal dependencies. The existing methods employ DCNN-based feature representation models [49, 88, 89], which may not adequately capture complex pathological ECG features and their temporal dependencies associated with the CHF syndrome. Moreover, these deeper networks are hard to optimize because of the vanishing gradient problem, which may degrade the diagnosis performance. Also, most existing DL methods lack "diagnostic transparency," i.e., discriminative input ECG features that led to the diagnosis are unknown. It limits the applicability of such "blackbox" methods in real-world clinical applications [49]. Therefore, it would be interesting to design a DL model that could effectively learn the complex pathological features and their temporal dependencies to improve the diagnosis performance and help visualize salient ECG features contributing the most to the diagnosis to enhance the model transparency.

- AF is the most common heart rhythm disorder that can progressively worsen over time. AF can be intermittent and asymptomatic; as a result, it often remains undiagnosed until it causes a life-threatening condition, such as stroke. It is thus a disorder that would benefit from long-term continuous monitoring with Holter ECG devices. Analyzing AF burden (the percentage of the time patient is in AF rhythm) from these long-term ECG recordings can help improve the diagnosis and severity assessment of AF condition, thereby initiating appropriate stroke prevention strategies [134–136]. However, the precise AF burden estimation from the long-term recordings is often challenging due to frequent ectopic beats, i.e., AF-masquerading rhythms and different noise levels. Existing approaches focus on detecting the presence/absence of AF and do not quantify the disease severity through AF burden measure. Therefore, a robust and accurate AF burden estimation method that can effectively handle the above challenges of long-term recordings is of great clinical significance.

- The simultaneous classification of multiple cardiac ailments, such as MI, CD or BBBs, and HYP can improve the applicability of automated methods in real-world clinical applications. However, these diseases show subtle and similar pathological ECG manifestations with wide morphological variabilities (multi-scale), challenging their reliable classification [120]. In addition, the co-occurrence of two or more cardiac disorders in patients further adds to the difficulty of developing effective automated ECG classification methods. In practice, clinicians often seek a second or more opinion for patients with confusing ECG biomarkers before making the final diagnosis. Thus, fusing suggestions from several experts can generate a reliable final decision. Most existing studies employ single expert or classifier-based approaches to classify cardiac disorders. However, it is argued that such approaches may not adequately represent the diverse pathological ECG manifestations of multiple cardiac ailments. We hypothesize that a mixture-of-experts-based fusion approach can effectively handle the disease variabilities by combining diagnosis decisions from several experts, which is expected to improve the multi-class and multi-label ECG classification performance.

1.5 Major Contributions of the Work

Motivated by the issues mentioned above, this dissertation aims to develop DL-based automated ECG analysis methods to handle these challenges effectively. The significant contributions are as follows.

- Multi-lead diagnostic attention-based recurrent neural network (MLDA-RNN) model has been investigated for MI severity stages classification. The model systematically processes the 12-lead ECG signals using lead-specific RNN layers followed by intra- and inter-lead attention modules. Specifically, the lead-specific RNN layers effectively characterize the temporal dynamics of subtle pathological manifestations associated with the MI severity. The subsequent intra- and inter-lead attention modules automatically emphasize/alleviate the diagnostically relevant/redundant within- and across-lead information during the feature fusion to obtain the high-level discriminative feature representations for reliable classification. The experimental analysis shows that the unique MLDA-RNN model with RNNs and attention modules can effectively combine the subtle pathological manifestations of 12-lead ECG for improved MI severity stages classification. This systematic feature fusion also helps eliminate the diagnostic redundancy of 12-lead ECG and aids in better generalization performance. Additionally, the analysis of attention weights demonstrates the correlation of weights with the diagnostic relevance of focused ECG features, which often aligns with medical knowledge.

- An attention-based deep residual recurrent neural network (A-DRRNet) architecture has been explored to capture the complex pathological ECG features and their temporal dependencies (PQRST) associated with the CHF syndrome. The model employs multi-layered RNNs consisting of recurrent hidden states to encode the temporal dynamics of complex pathological ECG characteristics effectively. In addition, residual connections have been introduced between the RNNs layers to ease the training of the proposed deeper network. Finally, we incorporate an attention module to further improve feature learning by focusing on clinically relevant ECG information for efficient CHF detection.

The experimental results show that the A-DRRNet model can learn highly discriminative features for improved CHF diagnosis. Moreover, the analysis of attention weight heat maps significantly improves the diagnostic transparency of the proposed model, which makes it reliable for clinical applications.

- A multi-task deep convolutional neural network (MT-DCNN) model has been devised for robust and accurate AF burden estimation from the long-term ECG recordings while effectively handling the AF-masquerading rhythms and noise artifacts. The model consists of AF detection as a primary task and reconstruction of ECG sequence as an auxiliary task. Specifically, we have explored using a convolutional denoising autoencoder (CDAE) structure to design the auxiliary task using DCNNs. This task regularizes the model to learn robust feature representations for efficient AF detection, thereby aiding accurate AF burden estimation. The experimental study on four diverse Holter ECG datasets confirmed that the proposed model outperforms the baseline approaches. The proposed multi-task model also provides a robust AF burden estimation in the presence of frequent ectopic beats and different noise levels over the baseline methods. In addition, the analysis of AF burden demonstrates improved diagnosis and stroke-risk stratification for paroxysmal AF patients.

- Inspired by the mixture-of-experts-based fusion strategy [137, 138], we have investigated a novel multi-scale deep temporal convolutional neural network ensemble (MS-DTCE) model that extracts and combines diverse multi-scaled ECG features for effective multi-class ECG classification. Specifically, we have designed several temporal CNN-based expert classifiers using dilated and causal convolutional filters with different receptive fields to extract the scale-specific pathological ECG characteristics and generate local predictions. In addition, a temporal CNN-based gating network is designed to aggregate the local predictions of experts based on their competencies to generate the final diagnosis decisions. A new objective function is formulated for optimizing the experts and the gating network simultaneously in an end-to-end manner. The experimental results show that the proposed deep ensemble framework can significantly improve the multi-class ECG classification of several acute cardiac disorders. A similar deep multi-scale ensemble framework is also extended for the multi-label ECG classification. The experimental results demonstrate that the ensemble model generates improved multi-label ECG classification performance. In addition, further analysis of multi-label diagnosis decisions reveals that the proposed model can identify low- and high-risk patients assessed in terms of the number of co-occurring cardiac disorders.

1.6 Organization of the Thesis

The proposed investigations that are presented in the rest of the thesis are illustrated in Figure 1.15. It consists of two major blocks, i.e., the single-lead ECG-based diagnosis and the 12-lead ECG-based diagnosis. Under the single-lead ECG-based diagnosis, heart rhythm disorders like AF and CHF are analyzed. On the other hand, in the 12-lead ECG-based diagnosis, several acute cardiac disorders such as MI, HYP, CD, and ST/T changes are investigated. **Chapter 2** to **Chapter 5** present the major research contributions of the thesis. Specifically, **Chapter 2** proposes a novel multi-lead diagnostic attention-based recurrent neural network (MLDA-RNN) model for MI severity stages classification from 12-lead ECG signals. It also presents the quantitative and qualitative evaluation of the proposed model to demonstrate its effectiveness for the MI severity assessment. **Chapter 3** investigates the applicability of single-lead ECG heartbeats for CHF detection using an attention-based deep residual recurrent neural network (A-DRRNet) architecture. It also explores the analysis of model interpretability results for the proposed and existing methods. **Chapter 4** presents a robust multi-task deep CNN (MT-DCNN) framework for estimating AF burden from the single-lead long-term ambulatory ECG recordings. This chapter systematically compares the effectiveness of the proposed and baseline approaches at frequent ectopic beats and different noise levels. It also discusses AF burden analysis for stratifying AF patients at high- and low-risk for stroke. **Chapter 5** investigates two multi-scale deep ensemble frameworks for multi-class and multi-label 12-lead ECG classification. The first approach presents a mixture-of-experts-based multi-scale deep temporal CNN ensemble framework for effective multi-class ECG classification. Similarly, the second approach presents a binary relevance-based ensemble of several multi-scale deep temporal CNN classifiers for multi-label ECG classification. This chapter also explores the analysis of multi-label diagnosis decisions for identifying low- and high-risk patients. A summary of the thesis work and the future directions are discussed in **Chapter 6**.

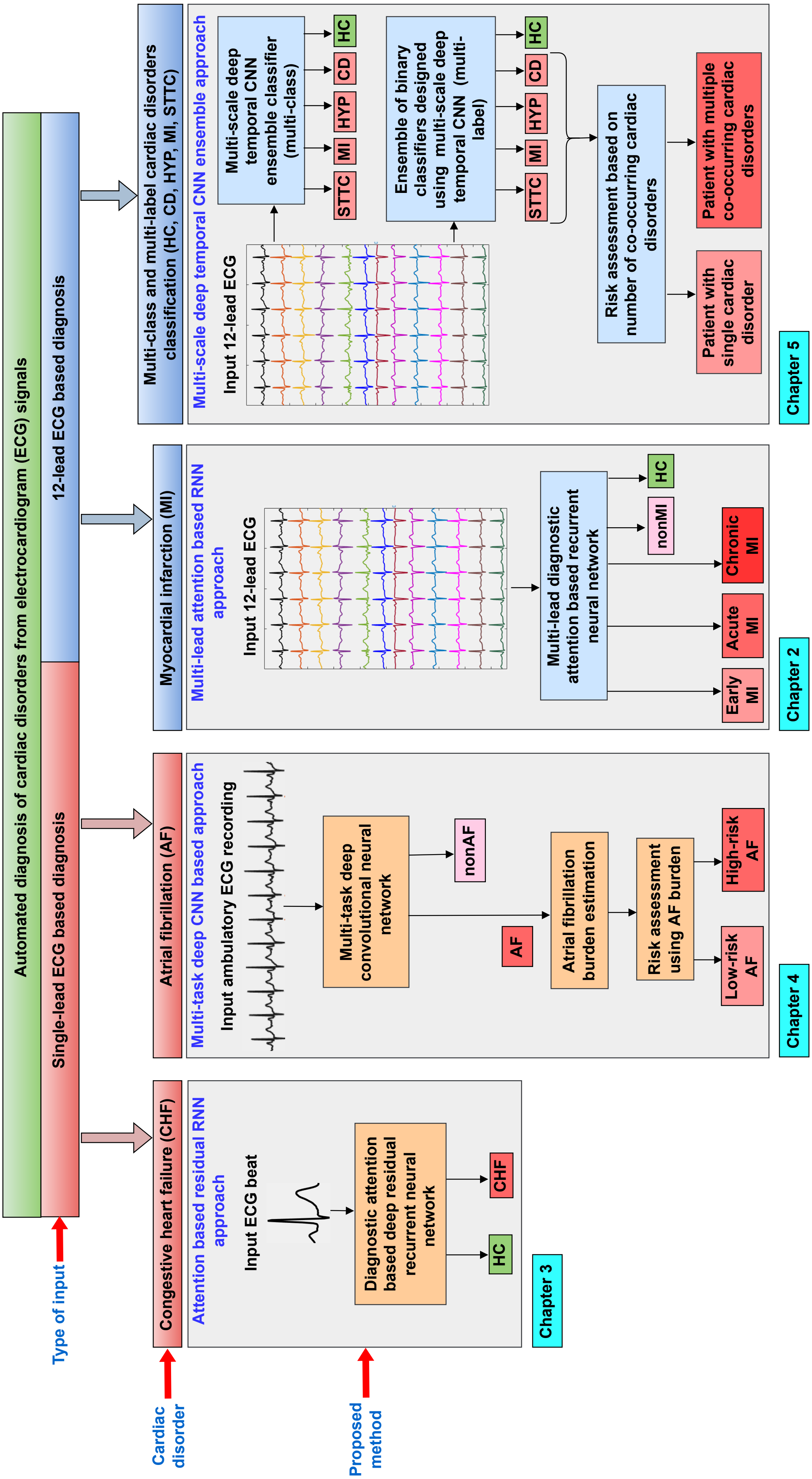


Figure 1.15: Graphical representation of the working chapters of this dissertation.

2

Myocardial Infarction Severity Stages Classification from 12-Lead ECG using Multi-Lead Attentional RNN

Contents

Myocardial infarction (MI), also known as heart attack, is one of the leading causes of death from CVDs globally [21]. As discussed in the previous chapter, MI is a progressive disorder that occurs due to the prolonged blockage of coronary arteries. Its progression can be broadly categorized into three stages, viz. early MI (occlusion persists for 40 min), acute MI (occlusion persists for 12-24 h), and chronic MI (occlusion persists for more than 24 h), based on the time from the onset of symptoms. In the early stages, MI can be reversible; however, if the occlusion persists without treatment for more than 24 h, the "myocardium at risk" becomes completely necrotic, causing irreversible myocardial damage with higher mortality rates. Thus, early diagnosis and severity assessment of MI using ECG are crucial in timely initiating appropriate treatment strategies that can avert further disease progression and reduce associated mortality.

The previous chapter discussed that during MI progression, the pathological ECG features such as T-wave peaking and inversion, ST-elevation, and pathological Q-waves at the leads monitoring the infarcted region undergo subtle and dynamic changes. For example, Figure 2.1 presents the association between MI progression and pathological ECG manifestations in lateral lead I for a patient before (Figure 2.1 (a)) and after LCx coronary artery occlusion. As can be seen, in the early stages of MI, the T-wave amplitude gradually increases (Figure 2.1 (b)) with slight ST-elevations (Figure 2.1 (c)). Similarly, because of the onset of necrosis in the acute MI stage, the appearance of pathological Q-waves, severe ST-segment elevations, and post-ischemic T-wave inversions can be observed on the ECG signals (Figure 2.1 (d)). Finally, the complete necrosis of the myocardium in the chronic MI stage results in distinct pathological Q-waves with ST-recovery and T-wave inversions (Figure 2.1 (e)). Based on the spatial location of the coronary occlusion, similar pathological ECG manifestations discussed above can appear in any of the ECG leads viewing the infarcted heart region. It is difficult to effectively represent the subtle pathological ECG changes and their dynamically varying temporal dependencies associated with the progressive MI condition. Moreover, the implicit intra- and inter-lead diagnostic redundancy of the 12-lead ECG often obscures the subtle yet clinically relevant information, which further adds to the difficulty of classifying MI severity stages.

Existing approaches consider MI as one broad category for the diagnosis and do not address its severity staging. These approaches mostly employ feature engineering [47, 48, 50, 51, 58, 60–64] and deep CNN-based [69, 71–78] models for the MI diagnosis from ECG signals. However, considering the progressive nature of the MI disease, the fixed handcrafted features or the deep CNNs may not effectively represent the temporal dependencies of subtle pathological ECG manifestations associated with the MI severity stages. Also, existing automated methods did not address the intra- and inter-lead diagnostic redundancies of the 12-lead ECG, which may limit them from learning discriminative feature representations for a reliable diagnosis.

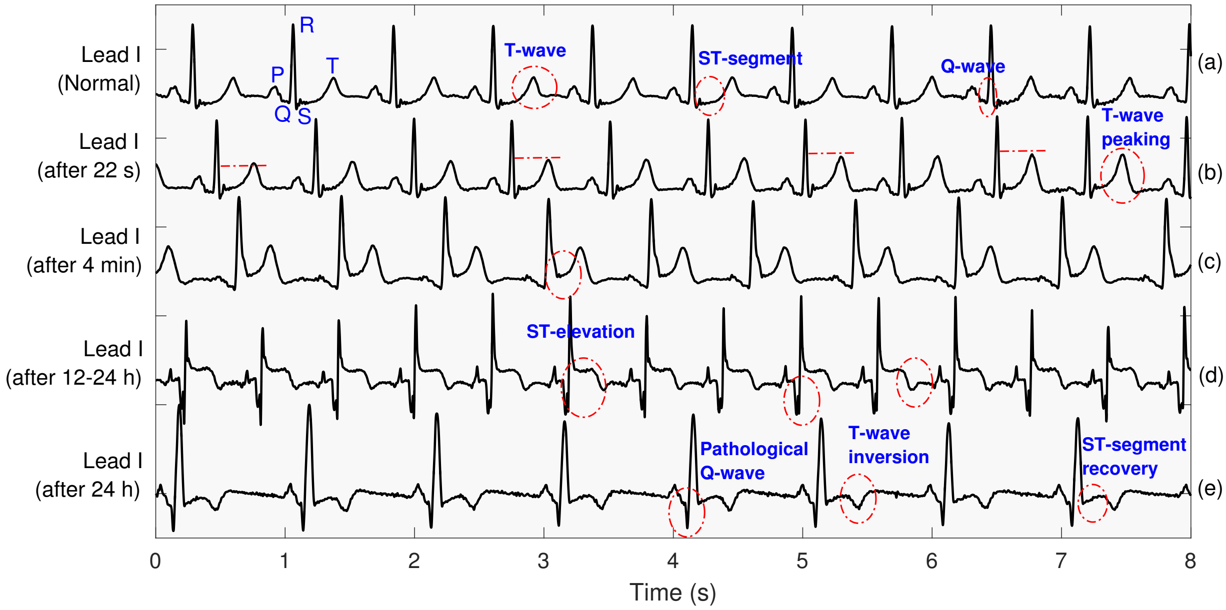


Figure 2.1: Illustrates the dynamic changes in pathological ECG characteristics during the MI progression at lateral lead I for an LCx coronary artery occlusion subject. (a) Pre-coronary occlusion normal ECG. (b) After 22 s of coronary occlusion (early MI). (c) After 4 min occlusion (early MI). (d) After 12-24 h of occlusion (acute MI). (e) After more than 24 h of occlusion (chronic MI).

The advancements in neural networks have led to the development of recurrent neural networks (RNNs). The RNNs can effectively model the temporal dependencies of sequential data and are shown to achieve state-of-the-art performance on various sequence modeling tasks such as speech recognition [139], time-series prediction [140], language modeling [140, 141], machine translation [141], and video tagging [140]. Considering the implicit temporal dependencies of the ECG waveform and their dynamic variations during the MI progression, RNNs have the potential to effectively model the 12-lead ECG variabilities of progressive MI severity stages. Moreover, in practice, during the 12-lead ECG evaluation, cardiologists carefully examine the salient ECG leads with distinct pathological ECG characteristics and fuse the clinical information from these leads to diagnose MI severity stages [142]. Recently, the development of deep attention networks has enabled the neural network to pay more attention to the useful features for efficient classification. The attention mechanism has been widely used in various fields such as neural machine translation (NMT) [141], image captioning [143], speech recognition [144], and brain-computer interference (BCI) [145]. Considering specific ECG leads (viewing the infarcted region) show distinct pathological manifestations such as T-wave inversions, ST-elevations, and pathological Q-waves, employing attention networks can aid in emphasizing this clinically relevant information of the 12-lead ECG for effective MI severity stages classification.

Inspired by the success of RNNs and attention networks, in this chapter, we propose a multi-lead diagnostic attention-based recurrent neural network (MLDA-RNN) for classifying MI severity stages from

12-lead ECG signals. The model systematically processes the 12-lead ECG using lead-specific RNN layers followed by intra- and inter-lead attention modules. Specifically, the lead-specific RNN layers effectively model the temporal dynamics of subtle pathological manifestations associated with the MI severity. The subsequent intra- and inter-lead attention modules automatically emphasize the diagnostically relevant within- and across-lead information of the 12-lead ECG during the feature fusion to obtain the discriminative feature representations for reliable MI staging. The rest of the chapter is organized as follows. Section 2.1 and 2.2 present the proposed framework and the experimental results for the MLDA-RNN method, respectively. The summary of the chapter is presented in section 2.3.

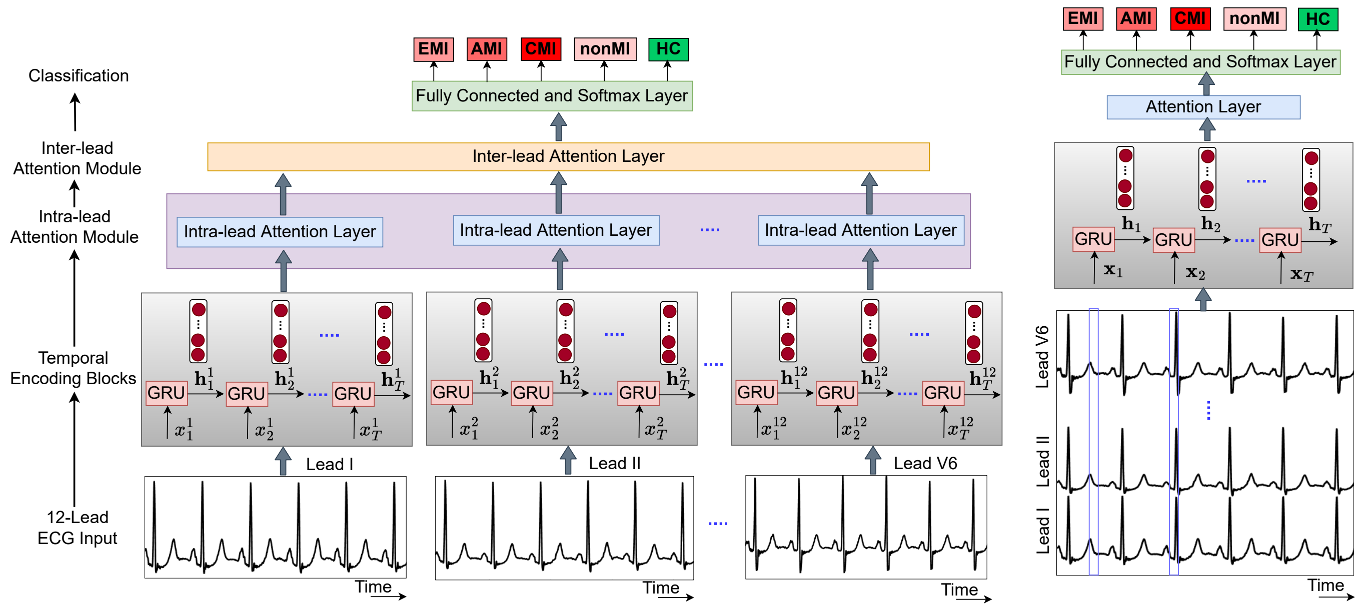
2.1 Multi-Lead Diagnostic Attention based RNN Approach for MI Staging

This section presents the proposed MLDA-RNN method for classifying MI severity stages from the 12-lead ECG signals. The block diagram of the method is shown in Figure 2.2 (a). It mainly consists of four processing steps. In the first step, the temporal encoding blocks with RNNs extract temporal features from each lead of the input 12-lead ECG. In the second and third steps, clinically relevant within lead temporal features and across lead diagnostic information are fused using intra- and inter-lead attention modules, respectively. The high-level attentively fused feature representation is fed to the classification block for classifying the input 12-lead ECG into one of the five categories, i.e., early MI (EMI), acute MI (AMI), chronic MI (CMI), non-MI, and HC in the final step. Here, the non-MI class accommodates patients with CVDs other than MI, and this inclusion help prevent misdiagnosis, thereby reducing over/under treatment [146]. The details of the different stages in the block diagram are discussed as follows.

2.1.1 Temporal Encoding with Recurrent Neural Networks

RNNs contain recurrent connections, in which the activation of the current hidden state depends on the previous time step [140]. In this way, the network captures the dynamic temporal behavior well. However, it is difficult to train the vanilla RNNs due to the exploding or vanishing gradient problem. The RNN variants, including long-short term memory (LSTM) and gated recurrent unit (GRU) structures, alleviate the vanishing gradient problem and can effectively learn the temporal dependencies [147]. Both contain sophisticated activations and gating functions to regulate the flow of information to maintain proper gradients till backward steps. Compared to LSTM, GRU has less trainable parameters with reduced complexity; thus, we chose GRU as the basic RNN unit to design temporal encoding blocks in the proposed MLDA-RNN.

Let $\mathbf{X} \in \mathbb{R}^{12\times T}$ be the 12-lead ECG sequence data and $\{\mathbf{x}_T^m \in \mathbb{R}^T | \mathbf{x}_T^m = [x_1^m, x_2^m, ..., x_t^m, ..., x_T^m]\}$, $m = \{1, 2, ..., 12\}$ be the m^{th} ECG lead. x_t^m represents m^{th}-lead t^{th} time step data and T is the sequence



(a) Proposed MLDA-RNN (Deep Fusion) (b) Proposed MLDA-RNN-E (Early Fusion)

Figure 2.2: (a) Block diagram of the proposed MLDA-RNN. The method process each lead separately using temporal encoding blocks with RNNs and intra-lead attention layers and fuse the 12-lead ECG feature representations using the inter-lead attention layer just before the classification, i.e., deep fusion. (b) Block diagram of the MLDA-RNN variant with early lead fusion using RNNs followed by attention layer and classification (MLDA-RNN-E).

length. The temporal encoding block at each ECG lead encodes the intra-lead or lead-specific temporal dependencies into a set of hidden vectors using GRU layers. Figure 2.3 (a) shows the typical GRU cell. It contains two important gates known as the reset gate ($\mathbf{r}_t^m$) and the update gate ($\mathbf{z}_t^m$) for controlling the flow of information inside the cell. $\mathbf{r}_t^m$ determines the part of the past information, which will be maintained in the current hidden state, $\mathbf{z}_t^m$ determines the part of new information, which will be added to the current hidden state. The activation of hidden layer for the given m^{th} ECG-lead at time step t is given by:

$$\mathbf{h}_t^m = (1 - \mathbf{z}_t^m) \odot \mathbf{h}_{t-1}^m + \mathbf{z}_t^m \odot \widetilde{\mathbf{h}}_t^m \tag{2.1}$$

where $\mathbf{h}_t^m$ and $\mathbf{h}_{t-1}^m$ represent the hidden state at time step t and $t-1$, respectively. $\odot$ denotes element-wise multiplication. $\mathbf{z}_t^m$ is the update gate and it can be derived as:

$$\mathbf{z}_t^m = \sigma(\mathbf{w}_{zi}^m x_t^m + \mathbf{W}_{zh}^m \mathbf{h}_{t-1}^m + \mathbf{b}_z^m) \tag{2.2}$$

where $\sigma(\cdot)$ is the Sigmoid activation function, $\mathbf{b}_z^m$ is a bias vector, $\mathbf{w}_{zi}^m$ and $\mathbf{W}_{zh}^m$ represent the input-update and update-hidden weights, respectively. The candidate hidden state $\widetilde{\mathbf{h}}_t^m$ can be computed as:

$$\widetilde{\mathbf{h}}_t^m = tanh(\mathbf{w}_{ci}^m x_t^m + \mathbf{W}_{cr}^m(\mathbf{r}_t^m \odot \mathbf{h}_{t-1}^m) + \mathbf{b}_c^m) \tag{2.3}$$

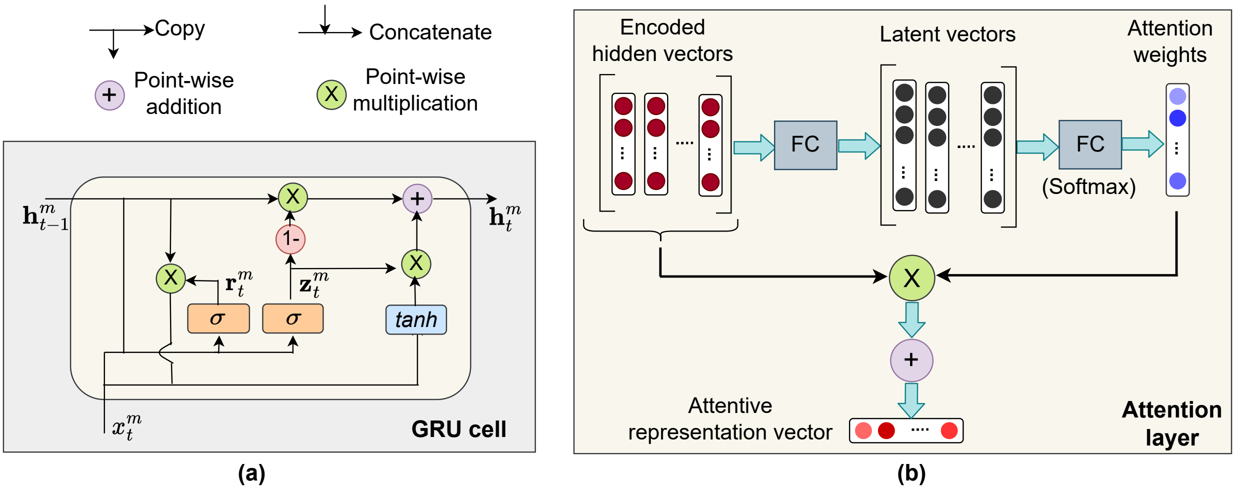


Figure 2.3: (a) A typical GRU cell architecture. (b) Self-attention layer architecture.

where $tanh$ is hyperbolic tangent activation function, $\mathbf{b}_c^m$ is bias vector, $\mathbf{w}_{ci}^m$ and $\mathbf{W}_{cr}^m$ denote the input-candidate, reset-candidate weights, respectively. Finally, the reset gate vector $\mathbf{r}_t^m$ can be given by:

$$\mathbf{r}_t^m = \sigma(\mathbf{w}_{ri}^m x_t^m + \mathbf{W}_{rh}^m \mathbf{h}_{t-1}^m + \mathbf{b}_r^m) \tag{2.4}$$

where $\mathbf{b}_r^m$ is bias vector, $\mathbf{w}_{ri}^m$ and $\mathbf{W}_{rh}^m$ denotes the input-reset, reset-hidden weights, respectively.

Finally, for the m^{th} ECG-lead $\{\mathbf{x}_T^m \in \mathbb{R}^T | \mathbf{x}_T^m = [x_1^m, x_2^m, ..., x_t^m, ..., x_T^m]\}$, its GRU encoded output hidden vectors are given as $\{\mathbf{H}_m \in \mathbb{R}^{d_h \times T} | \mathbf{H}_m = [\mathbf{h}_1^m, \mathbf{h}_2^m, ..., \mathbf{h}_t^m, ..., \mathbf{h}_T^m]\}$. Here, $\mathbf{h}_t^m \in \mathbb{R}^{d_h}$ represents m^{th} ECG-lead hidden vector at time step t, consisting of information about the input sequence up to x_t, and d_h denotes the size of the hidden vector and is known as GRU memory.

2.1.2 Intra-Lead Attention Module

As discussed previously, the ECG leads show intra-lead diagnostic redundancy, which often obscures the subtle pathological ECG manifestations such as T-wave peaking and inversion, ST-elevations, and pathological Q-waves, thereby limiting the classification performance of MI staging. Recently, the attention models have shown promising results in various recognition tasks by focusing on the salient features during the classification. In this study, an intra-lead attention layer is employed on top of the temporal encoding block of each ECG lead. These attention layers automatically discover the clinically relevant hidden vectors corresponding to the pathological ECG features and aggregate them with appropriate attention weights to improve the classification performance. The design of the attention module is as follows.

Each attention layer accepts the encoded hidden vectors from their respective temporal encoding block and generates soft attention weights for each of the encoded vectors (see Figure 2.3 (b)). It is expected that the attention weights are high for the vectors that correspond to the pathological ECG manifestations than the others. To achieve this, first, each encoded hidden vector $\mathbf{h}_t^m$ is non-linearly transformed into a latent space using a multi-layer perceptron (MLP) or fully connected (FC) layer as:

$$\mathbf{u}_t^m = tanh(\mathbf{W}_{intra}^m \mathbf{h}_t^m + \mathbf{b}_{intra}^m) \tag{2.5}$$

where $\mathbf{W}_{intra}^m \in \mathbb{R}^{d_a \times d_h}$ and $\mathbf{b}_{intra}^m \in \mathbb{R}^{d_a}$ are the weight matrix and the bias vector, respectively, of a MLP with latent size d_a. $\mathbf{u}_t^m \in \mathbb{R}^{d_a}$ denotes the latent vector of $\mathbf{h}_t^m$. The latent vectors ($\mathbf{u}_t^m$) are fed to another MLP/FC with Softmax activation, to obtain the weights of importance for each hidden vector $\mathbf{h}_t^m$ as:

$$\alpha_t^m = \frac{exp\left((\mathbf{u}_t^m)^\intercal \mathbf{u}_{intra}^m + b_{intra}^m\right)}{\sum_{t=1}^{T} exp\left((\mathbf{u}_t^m)^\intercal \mathbf{u}_{intra}^m + b_{intra}^m\right)} \tag{2.6}$$

where $\mathbf{u}_{intra}^m \in \mathbb{R}^{d_a}$ and $b_{intra}^m \in \mathbb{R}^1$ are the trainable parameters of the network and T is the transpose operator. The dot-product similarity between $\mathbf{u}_t^m$ and $\mathbf{u}_{intra}^m$ provides intra-lead attention weight α_t^m, $t = \{1, 2, ..., T\}$ for the m^{th} ECG lead. The $Softmax(\cdot)$ function guarantees that all the computed weights sum to one, and these weights will be relatively high for the hidden vectors corresponding to ECG features which are clinically more relevant for classification than others. Finally, the encoded vectors of each ECG lead, i.e., $\mathbf{h}_t^m$, $t = \{1, 2, ..., T\}$ are aggregated through weighted summation of the $\mathbf{h}_t^m$ and α_t^m, $t = \{1, 2, ..., T\}$ to obtain the within-lead attentive representation (WLAR) vector given as:

$$\mathbf{a}_m = \sum_{t=1}^{T} \alpha_t^m \mathbf{h}_t^m \tag{2.7}$$

where $\mathbf{a}_m \in \mathbb{R}^{d_h}$ denotes the WLAR vector for m^{th} ECG lead. Twelve such vectors (from 12-leads) are used to construct the within-lead attentive matrix $\mathbf{A} \in \mathbb{R}^{d_h \times 12}$ and is fed to the inter-lead attention module.

2.1.3 Inter-Lead Attention Module

As discussed before, based on the spatial location of the coronary occlusion and the area of infarction, specific leads of 12-lead ECG show the pathological manifestations, with others being clinically redundant. Thus, after summarizing the within-lead relevant ECG features, identifying the salient ECG leads and emphasizing them with suitable weights during the classification can help improve the performance of MI staging. To perform this task, we have employed an inter-lead attention module following the intra-lead attention layers. The intra-lead attention module aggregates the WLAR vectors corresponding to the 12-lead ECGs based on their importance in improving the overall classification. This summarization gives a high-level across-lead attentive representation (ALAR) vector and is computed using a similar formulation as intra-lead attention (Figure 2.3 (b)) given by:

$$\mathbf{v}_m = tanh(\mathbf{W}_{inter}\mathbf{a}_m + \mathbf{b}_{inter}) \tag{2.8}$$

$$\beta_m = \frac{exp\big((\mathbf{v}_m)^\intercal \mathbf{v}_{inter} + b_{inter}\big)}{\sum_{m=1}^{12} exp\big((\mathbf{v}_m)^\intercal \mathbf{v}_{inter} + b_{inter}\big)} \tag{2.9}$$

$$\mathbf{a}_f = \sum_{m=1}^{12} \beta_m \mathbf{a}_m \tag{2.10}$$

where $\mathbf{a}_f \in \mathbb{R}^{d_h}$ is the final ALAR vector with emphasis given to salient ECG leads relevant for classification thereby reducing inter-lead redundancy. $\mathbf{v}_m \in \mathbb{R}^{d_a}$ denotes the latent vector and β_m, $m = \{1, 2, ..., 12\}$ is the m^{th} ECG-lead attention weight. $\mathbf{W}_{inter} \in \mathbb{R}^{d_a \times d_h}$, $\mathbf{b}_{inter} \in \mathbb{R}^{d_a}$, $\mathbf{v}_{inter} \in \mathbb{R}^{d_a}$ and $b_{inter} \in \mathbb{R}^1$ are the trainable parameters of the network.

2.1.4 Classification Module

In the end, the compact representation vector $\mathbf{a}_f$ is fed to the output FC layer with Softmax activation (FC+Softmax) to obtain the class probabilities for the input 12-lead ECG $\mathbf{X}$. The probability distribution of the output categories obtained at the FC+Softmax layer is given as:

$$p(c|\mathbf{X}) = Softmax(\mathbf{W}_o\mathbf{a}_f + \mathbf{b}_o) \tag{2.11}$$

where $p(c|\mathbf{X})$ represents the probability of $\mathbf{X}$ belongs to class c, $c = \{1, 2, ...C\}$, C is the number of output categories, $\mathbf{W}_o \in \mathbb{R}^{C \times d_h}$ and $\mathbf{b}_o \in \mathbb{R}^C$ are weights and biases of the output layer. The final predicted class label for the input ECG $\mathbf{X}$ is given as:

$$class(\mathbf{X}) = \underset{c=1,2,...,C}{\arg\max}\; p(c|\mathbf{X}) \tag{2.12}$$

Training loss optimization: In general, the categorical cross-entropy loss between the probabilistic output and one hot encoded true label for a set of training samples is minimized to learn the network parameters of the model. Mathematically, the categorical cross-entropy loss ($\mathcal{L}$) is formulated as [70]:

$$\mathcal{L} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C}\mathbf{I}(y_n = c) log\big(p(c|\mathbf{X}_n)\big) \tag{2.13}$$

where $(\mathbf{X}_n, y_n)$, $n = \{1, 2, ..., N\}$ are the training examples, N is the number of training samples and $\mathbf{I}(\cdot)$ is an indicator function which is equal to one if y_n equals to c, otherwise equal to zero. The real-world clinical ECG datasets are often imbalanced, i.e., some diseases occur less frequently than others. The data imbalance poses a challenge for predictive modeling as the learning process is often biased toward the majority class examples, so minority ones are not well modeled. To alleviate this problem, a cost-sensitive variant of the categorical cross-entropy loss function is employed to train the prediction/classifier model. This cost-sensitive loss function penalizes the class errors differently. Specifically, the minority class errors are assigned with a higher weight value than the majority class sample's errors. The cost-sensitive loss function ($\mathcal{L}_w$) can be formulated as [70]:

$$\mathcal{L}_w = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C} w_c \mathbf{I}(y_n = c) log\big(p(c|\mathbf{X}_n)\big) \tag{2.14}$$

where w_c is the weight assigned to the classification errors of samples from class c and it is computed as the number of samples in the majority class to the number samples in the class c, i.e., the weights are inversely proportional to the class distribution. Thus, the cost-sensitive loss function assigns a higher weight to the training samples of the class with low sample size, thereby improving learning for these classes during the model training.

MLDA-RNN model with early ECG-lead fusion: Till now, we have discussed the deep fusion version of the MLDA-RNN model (see Figure 2.2 (a)), where each ECG lead is processed separately using temporal encoding blocks with RNNs followed by intra-lead attention layers. The 12-lead feature representations from these layers are fused using the inter-lead attention module just before the classification. Figure 2.2 (b) shows the early fusion variant of the MLDA-RNN named MLDA-RNN-E. As can be seen, instead of processing each ECG lead separately, the MLDA-RNN-E combines the temporal information across the 12-leads together at the first RNN layer itself, followed by the attention layer and classification module. In this study, the proposed two types of 12-lead ECG fusion using RNNs and attention layers, i.e., MLDA-RNN and MLDA-RNN-E, are evaluated to verify their effectiveness for MI severity stages classification.

2.2 Experimental Results and Discussion

This section discusses the details of the clinical ECG dataset and the evaluation scheme employed. This section also presents the network optimization details, experimental evaluation, and the analysis of the proposed MLDA-RNN model, followed by the performance comparison with the existing methods.

2.2.1 Clinical ECG Database

The proposed method is evaluated on two publicly available 12-lead ECG PhysioNet databases [148]: the STAFF III database and the Physikalisch-Technische Bundesanstalt (PTB) diagnostic database. In this study, the input ECG length is chosen to be 4 s as progressive diseases like MI require an ample number of consecutive beats for analysis [60]. It allows us to capture the intra- and inter-beat dependencies well and

Table 2.1: Details of the ECG databases used for evaluation.

Class	Database	# Subjects	(# Male; # Female)	# 12-lead 4 s segments
EMI	STAFF III [148]	52	38; 14	3610
AMI	PTB [148]	55	42; 13	2110
CMI	PTB [148]	109	83; 26	3615
non-MI	PTB [148]	54*	35; 19	1905
HC	PTB [148]	52	39; 13	3020
Total	-	**322**	**227; 95**	**14260**

#: Number of; *: non-MI patients distribution (mimicking MI: 16 and others: 38)

aid in improving the performance (see Figure 2.5). The specific details of the databases are given below.

2.2.1.1 STAFF III Database

The ECG data for EMI patients are selected from this database. STAFF III provides induced early progression of MI (EMI) data from patients receiving elective percutaneous transluminal angiography in one of the major coronary arteries at the Charleston Area Medical Center in West Virginia. The mean balloon inflation time is 4 min 23 s, and this data helps analyze the early ECG manifestations of MI. In this study, we consider the first occlusion ECG recordings of 52 subjects, where each recording consists of nine leads (V1-V6, I, II, and III), sampled at 1000 Hz with an amplitude resolution of 0.6 μV. The remaining three leads (aVR, aVL, and aVF) are computed from limb leads to form the standard 12-lead ECG.

2.2.1.2 PTB Diagnostic Database

The ECG data for the AMI, CMI, non-MI, and HCs are extracted from this database. The PTB database consists of 12-lead ECG recordings from 290 individuals, and each record is sampled at 1000 Hz with an amplitude resolution of over $\pm$16.38 mV range. From the patient annotations, based on the infarct age and interventions, we selected data for AMI and CMI patients with infarct age less than 12-24 h and more than 24 h, respectively. The non-MI data consists of ECG records from other cardiac disorders except MI, such as HYP, BBBs, myocarditis, dysrhythmia, and valvular heart diseases. It is worth emphasizing that HYP, BBBs, and myocarditis are known as mimicking MI diseases due to their MI-like ECG features, which pose a challenge for reliable MI staging.

This study excludes prior MI patients from both databases to obtain good representative data for AMI and CMI. Moreover, patients with inappropriate annotations are also excluded from the analysis. The multi-lead signal quality estimation method [149] is employed to exclude clinically unsuitable 12-lead ECGs.

In order to reduce the computational burden, all the ECG signals are downsampled to 125 Hz. Table 2.1 presents the details of the selected ECG datasets across the five classes under analysis.

2.2.2 Evaluation Scheme and Performance Measures

Two evaluation schemes, namely class-oriented and patient-independent, are commonly used for the ECG-based disease classification tasks [150]. In the class-oriented scheme, ECG signals from training and test sets may belong to the same patient, which results in an overestimation of the classification performance. The well-trained classifier may fail to predict the disease labels for the ECG signals from an unseen individual. The scheme is not applicable in practice because of the significant variation in ECG characteristics among different subjects. On the other hand, a patient-independent evaluation scheme ensures no overlap of patient's data while dividing train and test datasets. The classifiers designed and evaluated using this scheme perform more realistically. Therefore, we employ a patient-independent scheme for evaluation. In addition, a 5-fold cross-validation (CV) protocol (in a patient-independent manner) is considered for the extracted dataset. Cross-validation ensures generalizable results and prevents overfitting or bias to specific subjects of the databases. The extracted data (see Table 2.1) exhibits an imbalance with the varied class distribution. Therefore, the class-weighted categorical cross-entropy loss is used for training classifiers.

Table 2.2: A multi-class confusion matrix.

		Predicted output			
		1	2	...	C
Actual output	1	u_{11}	u_{12}	...	u_{1C}
	2	u_{21}	u_{22}	...	u_{2C}
	.	.	.	...	.
	.	.	.	...	.
	.	.	.	...	.
	C	u_{C1}	u_{C2}	...	u_{CC}

The classification performance is evaluated using the class-wise and the overall measures derived from the confusion matrix. Table 2.2 shows the representation of a multi-class confusion matrix with entries u_{ab} representing the number of samples from the test set originally belonging to class a and classified as class b where $a, b \in \{1, 2, ..., C\}$. The class-wise measures include the sensitivity (Se), the specificity (Sp) and the accuracy (Acc). The mathematical formulation of these measures are given in Eq. 2.15 where Se_a, Sp_a and the Acc_a are the class-wise sensitivity, specificity and accuracy for class $a, a \in \{1, 2, ..., C\}$. The overall measures used are the overall sensitivity (OS), the overall specificity (OSp) and the overall accuracy (OA) and are computed as shown in Eq. 2.16.

$$Se_a = \frac{u_{aa}}{\sum_{e=1}^{C} u_{ae}} \qquad Sp_a = \frac{\sum_{k\neq a}^{C}\sum_{e\neq a}^{C} u_{ke}}{\sum_{k\neq a}^{C}\sum_{e\neq a}^{C} u_{ke} + \sum_{k\neq a}^{C} u_{ka}} \qquad Acc_a = \frac{u_{aa} + \sum_{k\neq a}^{C}\sum_{e\neq a}^{C} u_{ke}}{\sum_{k=1}^{C}\sum_{e=1}^{C} u_{ke}} \tag{2.15}$$

$$OS = \frac{1}{C}\sum_{a=1}^{C} Se_a \qquad OSp = \frac{1}{C}\sum_{a=1}^{C} Sp_a \qquad OA = \frac{\sum_{k=a}^{C} u_{ke}}{\sum_{k=1}^{C}\sum_{e=1}^{C} u_{ke}} \tag{2.16}$$

The area under the curve (AUC) and the Cohen's Kappa score [151] are also employed for the performance analysis.

2.2.3 Network Parameters

Each lead ECG with 500 samples ($T = 500$), are encoded with single layer RNN of size 32 ($d_h = 32$) to obtain hidden vectors. These vectors are fed to the respective intra-lead attention module with latent space dimension 64 ($d_a = 64$) to form WLAR vectors ($\mathbf{a}_m$) . The WLAR vectors obtained by emphasizing within lead diagnostic information, thereby reducing intra-lead redundancies. Later, these WLAR vectors are fed to inter-lead attention module with latent space size 64 ($d_a = 64$) to generate ALAR vector ($\mathbf{a}_f$). This vector attends the most informative leads, thereby reducing inter-lead redundancy. This reduction of intra- and inter-lead redundancies will improve the classification by generating most discriminative high-level representations. Finally, the 32-dimensional ALAR vector is fed to the FC+Softmax layer to classify the MI severity stages. To improve the speed of convergence and performance of the model, we employed batch normalization (BN) layers at the intra- and inter-lead attention layers. A dropout layer, with a drop rate of 0.3 is used at the output layer to improve the model generalization performance.

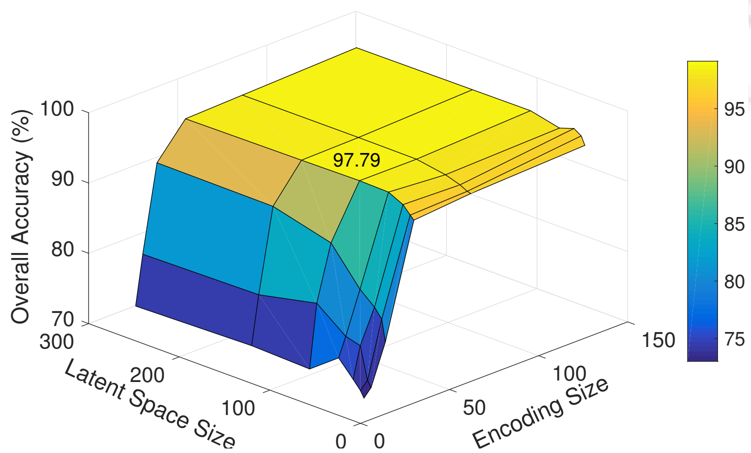


Figure 2.4: Performance variation of the MLDA-RNN with different encoding size (d_h) and latent space size (d_a).

The hyperparameters, RNN encoding size (d_h) and latent-space size (d_a) used in the intra- and inter-lead attention module are optimized using grid search technique. To verify the effect of these parameters on the classification performance, we fix the number of RNN layers to one and select d_h from the set $\{4, 8, ..., 128\}$. Then we fix d_h and choose d_a from the set $\{4, 8, ..., 256\}$. From Figure 2.4, it can be observed that the proposed method achieves better OA when $d_h \geq$ 32 and $d_a \geq$ 64 than the other values. The OA is poor at lower values of d_h due to not having sufficient memory. Once enough memory ($\geq$ 32) has been employed, OA increases to 97.79 $\%$ at d_a equal to 64. It can also be seen that, when we increase d_h and d_a beyond 64 and 128, due to an increase in complexity of the network for the available dataset, the network starts over-fitting. Thus, the grid search analysis results in the optimized d_h and d_a values as 32 and 64, respectively. Similarly, for the early fusion variant of MLDA-RNN, i.e., MLDA-RNN-E, the optimized d_h and d_a values are 64 and 128, respectively.

The network is developed using Keras with a TensorFlow backend. The multi-class loss is calculated using Eq. 2.14 and the model is trained using backpropagation through time (BPTT) [140]. The model is trained with an Adam optimizer with a learning rate of 0.001 on mini-batches of size 32 for 200 epochs.

2.2.4 Evaluation of the Proposed MLDA-RNN and MLDA-RNN-E Methods

This section presents the performance comparison of the proposed MLDA-RNN method and the MLDA-RNN-E method. Table 2.3 shows the classification results of these methods in terms of the class-wise and overall performance measures. As can be seen, the MLDA-RNN method provides improved classification results compared to the MLDA-RNN-E approach. The first observation from the results is that processing 12-lead ECG together with the early fusion (MLDA-RNN-E) may hamper the lead-specific information, which is crucial for the reliable diagnosis of MI staging, thereby resulting in sub-optimal performance (see Table 2.3). However, the proposed MLDA-RNN method processes the 12-lead ECG systematically using lead-specific RNNs and intra-lead attention layers and attentively aggregates the 12-lead feature representations using the inter-lead attention module. This deep fusion of the 12-lead ECG using MLDA-RNN aid in extracting discriminative feature representations, thereby improving the classification performance (see Table 2.3). Specifically, the proposed MLDA-RNN method achieves an OS, OSp and OA of 97.60$\%$, 99.43$\%$ and 97.79$\%$, respectively. The method also shows promising results for confusing classes such as AMI and CMI with Se of 94.78% and 97.64%, respectively. Similarly, class-wise Sp and Acc for all the classes are well above 98% indicates that the proposed model is near accurate in identifying true negatives and true positives (see Table 2.4). The improved performance of the MLDA-RNN method can be attributed to the lead-specific temporal encoding using RNNs and systematic feature fusion using attention modules that effectively capture the within-lead and across-lead subtle diagnostic information of the 12-lead ECG for efficient classification.

Table 2.3: Performance comparison of the MLDA-RNN-E (early fusion) and the MLDA-RNN (deep fusion) methods in terms of mean and standard deviation values of the class-wise and overall measures across 5-folds.

Method	**Class**	**Se**(%)	**Sp**(%)	**Acc**(%)	**OS**/**OSp**/**OA** (%)
MLDA-RNN-E	EMI	96.26 ± 1.22	98.59 ± 0.76	98.01 ± 0.83	
	AMI	84.83 ± 5.96	97.65 ± 0.98	95.75 ± 1.42	91.44 ± 3.88
	CMI	90.73 ± 3.72	96.33 ± 1.08	94.41 ± 2.54	98.03 ± 0.42
	non-MI	88.71 ± 5.11	99.31 ± 0.09	97.89 ± 0.31	91.84 ± 1.54
	HC	96.68 ± 1.18	98.26 ± 0.69	97.93 ± 1.02	
MLDA-RNN	EMI	98.34 ± 0.10	99.43 ± 0.08	99.15 ± 0.26	
	AMI	94.78 ± 0.88	99.62 ± 0.11	98.91 ± 0.29	97.60 ± 0.54
	CMI	97.64 ± 0.34	98.87 ± 0.06	98.56 ± 0.33	99.43 ± 0.08
	non-MI	97.90 ± 0.29	99.71 ± 0.06	99.47 ± 0.02	97.79 ± 0.47
	HC	99.33 ± 0.38	99.51 ± 0.14	99.47 ± 0.26	

Table 2.4: Confusion matrix of the proposed MLDA-RNN method across 5-folds.

		Predicted output				
		EMI	AMI	CMI	non-MI	HC
Actual output	EMI	3544	20	26	5	15
	AMI	9	2001	78	22	0
	CMI	11	24	3528	10	42
	non-MI	33	0	12	1860	0
	HC	11	0	10	1	2998

2.2.5 Effectiveness of the Proposed MLDA-RNN Method

In this section, we analyze the effectiveness of the proposed MLDA-RNN model from two perspectives; (i) the significance of the intra- and inter-lead attention modules and (ii) the visualization of intra- and inter-lead attention weights to infer the model interpretability results, i.e., ECG leads and ECG features focused by the model during the classification process.

2.2.5.1 Significance of Intra- and Inter-lead Attention Modules

Table 2.5 presents the performance comparison of the proposed MLDA-RNN model with two variants, the RNN-I model with no attention modules and the RNN-II model with only intra-lead attention. Specifically, RNN-I uses the last hidden vector from the RNN-based temporal encoding blocks of each ECG lead and averages them across 12-leads before feeding it to the classifier (Figure 2.2 (a)). On the other hand, the RNN-II with intra-lead attention module combines within-lead relevant hidden vectors to form WLAR, and these vectors are averaged across all the 12-leads and fed to the classifier (Figure 2.2 (a)). From Table 2.5,

Table 2.5: The performance comparison of the RNN-I model (no attention), the RNN-II model (only intra-lead attention), and the proposed MLDA-RNN (both intra- and inter-lead attention).

Method	**Class**	**Se**(%)	**Sp**(%)	**Acc**(%)	**OS/OSp/OA** (%)
RNN-I (no attention)	EMI	94.04	94.92	94.70	
	AMI	**9.47**	98.97	85.72	**71.67**
	CMI	**82.84**	82.94	82.92	93.80
	non-MI	**81.62**	97.77	95.61	**76.26**
	HC	90.39	94.38	93.54	
RNN-II (only intra-lead attention)	EMI	95.42	99.20	98.24	
	AMI	**77.01**	96.29	93.43	**90.52**
	CMI	**87.41**	95.01	93.08	97.74
	non-MI	**95.27**	99.43	98.87	**91.08**
	HC	97.50	98.79	98.52	
Proposed (intra- and inter-lead attention)	EMI	98.34	99.43	99.15	
	AMI	**94.78**	99.62	98.91	**97.60**
	CMI	**97.64**	98.87	98.56	99.43
	non-MI	**97.90**	99.71	99.47	**97.79**
	HC	99.33	99.51	99.47	

it can be observed that the proposed model outperforms RNN-I and RNN-II models in terms of class-wise and overall measures. The RNN-I model shows poor Se for CMI, non-MI, and HC classes and very poor for the AMI class. The intra-lead attention module in RNN-II helps considerably improve performance for all the classes. Still, the AMI and CMI classes are not modeled well, which leads to their misdiagnosis. However, the proposed method with intra- and inter-lead attention modules helps improve performance for all the classes. Specifically, as can be observed in Table 2.5, from RNN-I to the proposed model, there is a significant improvement in the OS, OSp, and OA of about 26$\%$, 5$\%$, and 21$\%$, respectively. Therefore, the intra- and inter-lead attention modules in the MLDA-RNN model contribute individually to improving the performance by attending to relevant within- and across-leads ECG information.

We have also analyzed the effect of different input ECG lengths on the overall classification performance. Figure 2.5 shows the variation of OA for the RNN-I, the RNN-II, and the proposed method across different ECG input lengths. As can be seen, the RNN-I shows poor results due to too condensed last hidden vectors used for classification. This problem becomes severe for long-duration ECGs. However, incorporating intra- and inter-lead attention modules in the MLDA-RNN minimizes this problem by selecting clinically relevant information from 12-lead ECGs. It can be seen from Figure 2.5 that the proposed model outperforms RNN-I and RNN-II, irrespective of ECG length. Specifically, the proposed model achieves the best performance at an input length of 4 s. Therefore, we choose 4 s ECG excerpts for the MI severity staging in the proposed framework. The results also suggest that the models with attention layers (RNN-II and proposed) effectively analyze the long-term ECG segments over the model with no attention layers (RNN-I).

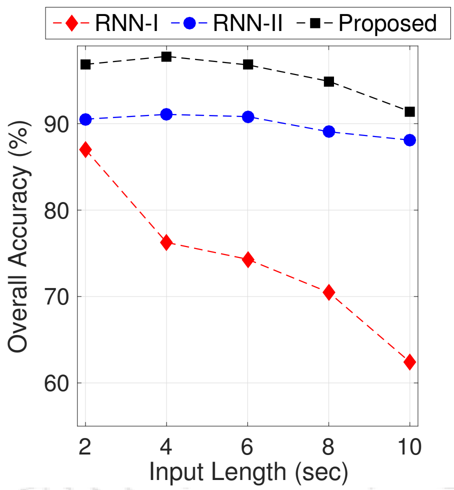


Figure 2.5: Overall accuracy comparison for RNN-I, RNN-II, and proposed model over different input ECG lengths.

2.2.5.2 Anlysis of Model Interpretability using Intra- and Inter-Lead Attention Weights

Model explainability is equally crucial as diagnostic accuracy in healthcare applications, which most of the existing DL-based methods lack [70, 119]. The certainty must exist that the disease diagnosis is made based on the real ECG characteristics that reflect pathology rather than systematic biases in the dataset. From this perspective, the proposed model's intra- and inter-lead attention weights are visualized and analyzed in the present section.

Analysis of inter-lead attention weights: Figure 2.6 shows the distribution of inter-lead attention weights learned by the proposed MLDA-RNN model for 12-lead ECG on one of the validation folds. As can be seen, the proposed model provides different weights of importance for each ECG lead based on their diagnostic relevance or redundancy in classifying the classes under analysis. Specifically, among the six frontal-plane ECG leads, the leads I and aVL contribute most to the classification process. The physiological reason for the model to give more weight to these two leads is that they monitor the left side of the heart and are critical for diagnosing MI and non-MI patients (BBBs and HYP) [11]. Moreover, the ECG-leads I and aVL contain all the information needed to accurately derive the other four frontal plane or limb leads. Interestingly, the method also learned to focus on only these two ECG leads from the frontal-plane leads. Similarly, among the six chest leads (V1-V6), the model emphasizes the ECG leads V1, V3, and V5 for reliable classification (Figure 2.6). The leads V1, V3, and V5 are favored among the 12-leads by the physicians for identifying septal wall MI, anterior wall MI and lateral wall MI, respectively [11, 73]. In addition, the right-sided lead V1 and the left-sided lead V5 are popularly used in clinics to detect non-MI diseases such as BBBs (left and right) and left ventricular hypertrophy by recognizing the wide bizarre QRS-complexes or tall R-waves.

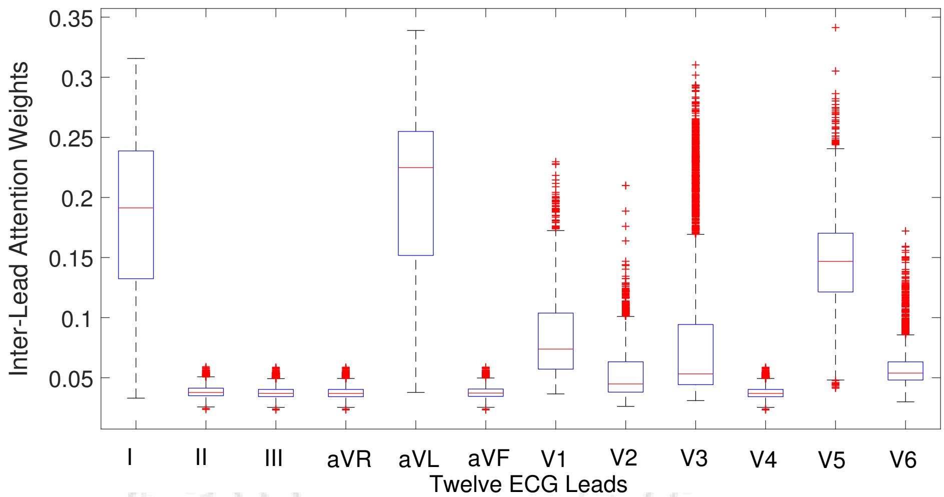


Figure 2.6: Boxplot of the inter-lead attention weights learned by the proposed MLDA-RNN model. The figure illustrates the importance of each ECG-lead during the classification process.

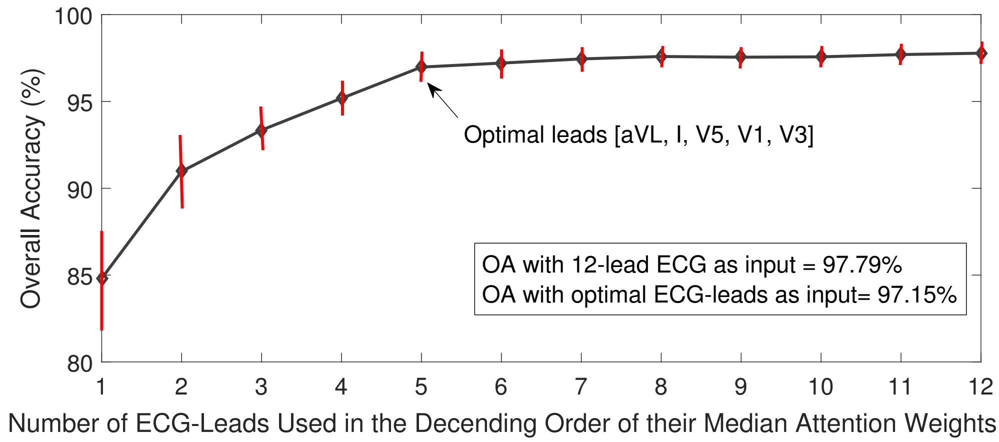


Figure 2.7: Optimal ECG-leads selection based on performance (mean $\pm$ standard deviation) at highly attentive ECG lead combination.

Figure 2.7 illustrates the effectiveness of inter-lead attention weights in identifying optimal ECG lead combinations for the classification. Using the proposed MLDA-RNN, the overall classification accuracy is computed at different ECG-lead combinations arranged in the descending order of their median attention weights. For evaluating the single-lead ECG performance, the MLDA-RNN architecture is modified as RNN layer$\rightarrow$intra-lead attention layer$\rightarrow$output classification layer. As seen from Figure 2.7, the ECG leads subset, i.e., aVL, I, V5, V1, and V3 combination, significantly improved ($p<0.05$) the classification performance over the single-lead and the inclusion of additional leads did not improve the performance significantly. Notably, the similar performance for 12-lead ECG and optimal ECG leads (Figure 2.7) demonstrate that the proposed MLDA-RNN model automatically identifies the diagnostically relevant and redundant ECG leads for improving classification performance.

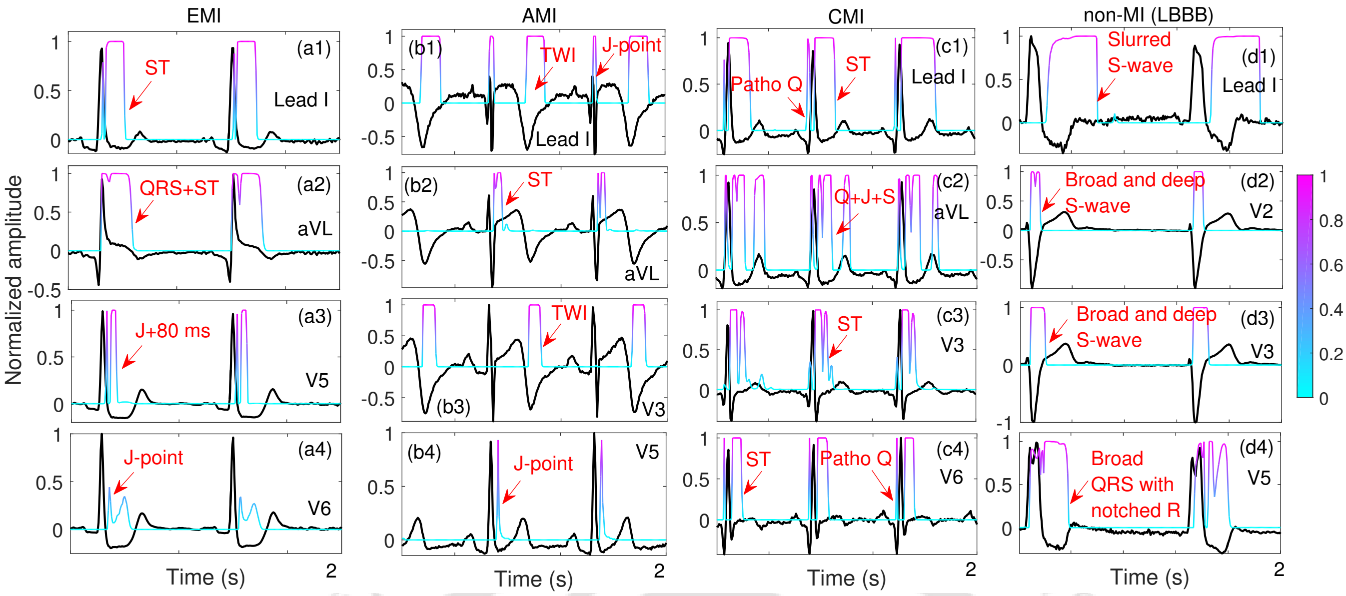


Figure 2.8: Illustrates the mapping of intra-lead attention weights corresponding to the frequently attended ECG leads (shown for 2 s) of (a1)-(a4) EMI, (b1)-(b4) AMI, (c1)-(c4) CMI and (d1)-(d4) LBBB patients, respectively. Colour bar indicates the intensity magnitude proportional to clinical relevance. (J+80msec: 80 msec after J-point; ST:ST-segment; Q+J+S: QRS, J-point and ST-segment; QRS+ST: QRS and ST-segment; TWI: T-wave inversion; Patho q: Pathological Q-wave).

Analysis of intra-lead attention weights: Figure 2.8 shows the visualization of typical intra-lead attention weights for the most focused ECG leads of four classes. The first observation from the figure is that the intra-lead attention layers provide different weights for different ECG features based on their relevance in diagnosing MI and non-MI pathologies. In clinical practice, physicians often examine ST-segment abnormalities (elevation or reciprocal depression), poor R-wave progression, T-waves inversion, and pathological Q-waves to detect MI progression [11]. In addition, J-point and J+80 msec points are also examined to understand MI. Similarly, broad and clumsy QRS-complexes in leads V5/V6/I and deep and broad S-waves in V2/V3 are used in clinics to detect the left BBB patients [11]. From Figure 2.8 (a1)-(a4) for EMI (with septal MI) and Figure 2.8 (b1)-(b4) for AMI patients (with anterior-lateral MI), the model provides more weight to the above discussed clinically relevant ECG features such as ST-segment, J-point, J+80 msec, T-wave, and QRS-abnormalities during the classification process. Interestingly, from Figure 2.8 (c1)-(c4) for CMI patients (with lateral MI), the pathological Q-waves in leads I and V6, along with the ST-segment, are focused by the model for the diagnosis. Similarly, from Figure 2.8 (d1)-(d4) for the non-MI patient (left BBB), broad QRS-complexes and deep S-waves are emphasized during the classification.

In summary, the analysis of inter-lead attention weights (Figure 2.6) and the intra-lead attention weights (Figure 2.8) demonstrates that the proposed model focuses on the clinically relevant ECG leads and features to make the reliable class prediction that often aligns with medical knowledge.

Table 2.6: Comparison of the proposed method against existing methods on MI detection using PTB Dataset

Method	#Classes	#Leads	Decision	Selected dataset	#Features	Classifier	Results	Evaluation	Riskstratification
[60]	2 (MI,HC)	12	Segment	MI:100 HC:100	Multiscale energy and eigen value features)	SVM	Se=93% Sp=99.00% Acc=96.00%	Class based	×
[64]	2 (MI,HC)	12	Segment	MI:100 HC:74	Empirical WT features	SVM	Se=100% Sp=99.95% Acc=99.97%	Class based	×
[61]	2 (MI,HC)	12	Beat	MI:28213 HC:5373	Wavelet packet transform based	SVM	Se=80.96% Sp=80.96% Acc=92.69%	Class based	×
[69]	2 (MI,HC)	1	Beat	MI:40182 HC:10546	End-to-end CNN (with R-peak detection)	Softmax	Se=93.71% Sp=92.83% Acc=93.53%	Class based	×
[72]	2 (Inferior MI,HC)	3	Segment	MI:3222 HC:3055	End-to-end CNN	Softmax	Se=85.33% Sp=84.04% Acc=84.54%	Patient-independent	×
[75]	2 (Anterior MI,HC)	3	Beat	MI:41087 HC:18640	End-to-end CNN (with R-peak detection)	Softmax	Se=85.33% Sp=84.09% Acc=84.54%	Class based	×
[73]	2 (Anterior MI,HC)	4	Beat	MI:34769 HC:34769	End-to-end CNN (with R-peak detection)	Softmax	Se=95.40% Sp=97.37% Acc=96.00%	Patient-independent	×
[71]	2 (MI,HC)	12	Beat	MI:485752 HC:125652	End-to-end CNN (with R-peak detection)	Softmax	Acc=99.78%	Class based	×
[77]	2 (MI,HC)	12	Beat	MI:48690 HC:10646	End-to-end MFB-CNN (with R-peak detection)	Softmax	Se=99.97% Sp=99.90% Acc=99.95%	Class based	×
[78]	2 (MI,HC)	12	Beat	MI:53712 HC:10638	End-to-end MFB-CBRNN (with R-peak detection)	Softmax	Se=94.42% Sp=86.29% Acc=93.08%	patient-independent	×
Proposed†	3 (MI, non-MI, HC)	12	Segment	MI:9335 non-MI:1905 HC:3020	End-to-end intra- and inter-lead attention based RNN model	Softmax	OS=98.63% OSp=99.27% OA=98.87%	Patient-independent	×
Proposed	5 (EMI,AMI, CMI,non-MI, HC)	12	Segment	EMI:3610 AMI:2110 CMI:3615 non-MI:1905 HC:3020	End-to-end intra- and inter-lead attention based RNN model	Softmax	OS=97.60% OSp=99.43% OA=97.79%	Patient-independent	√

#: Number of; and the selected dataset numbers correspond to beats or segments according to the type of decision.
†: The MI diagnosis results are generated by considering the outputs from EMI, AMI, and CMI classes as MI.

2.2.6 Performance Comparison with the Existing Methods

This section compares the proposed MLDA-RNN model with the existing methods. As can be seen from Table 2.6, all the existing methods [60, 61, 64, 69, 71–73, 75, 77, 78] have performed the binary classification of MI and HCs using ECG. However, in this study, we addressed the classification of MI severity stages which can help clinicians optimize various treatment strategies based on the disease progression. Secondly, some of the existing methods [69, 72, 73, 75] use subset of 12-lead ECG for MI/localized MI detection. It can be seen from Table 2.6 that these methods result in poor generalization, and it is evident from [72, 75] as average accuracy is below 85%. Moreover, the prior selection of specific leads from the 12-lead ECG makes these approaches unsuitable in real-world clinical settings [21], where patients can present with MI at any spatial location. The class-based evaluation scheme in [60, 61, 64, 69, 71, 75, 77] demonstrates overestimated performance because signals in the training and test sets could belong to the same patient. Thus, these methods may fail to predict diseases from ECG signals of unseen patients. Table 2.6 also highlights that the most existing methods require R-peak or beat detection as a pre-processing step. This increases the computational burden and may also be prone to erroneous diagnosis in the presence of noises [59]. The comparison methods in Table 2.6 can be broadly categorized into the ML- and DL-based methods. Although ML-based methods [60, 61, 64] achieved good results for MI diagnosis on smaller datasets, these methods rely on domain expert knowledge to engineer valuable features, which often faces the challenge of feature quality and robustness. The end-to-end DL approaches [69, 71–73, 75, 77, 78] mitigate these challenges and provide better generalization performance for the diagnosis on the larger and diverse ECG datasets. Existing methods only performed MI diagnosis and did not address MI staging. In order to compare the MI diagnosis performance of the proposed MI staging method with the existing approaches, we have combined the output decisions from EMI, AMI, and CMI classes as MI. The proposed method's three-class (MI, non-MI, and HC) classification performance is reported in Table 2.6. As can be seen, the proposed RNNs and attention-based model (MLDA-RNN) demonstrate improved MI diagnosis performance over existing DL methods, even in the presence of non-MI class patients and the patient-independent evaluation scheme.

In order to perform the qualitative comparison of the proposed MI staging method, we have implemented some recent DL-based models [72, 77, 78]. All these models are initially proposed to detect MI from HC subjects. However, they have been optimized for the current five-class ECG dataset. This experiment helps evaluate the feature representation ability of the proposed method and the recent DL models. The models considered for the comparison are; (i) deep CNN-based framework using lead II, III, and aVF ECG signals (3L-CNN) and (ii) two 12-lead ECG-based models, i.e., MFB-CNN [77] and MFB-CBRNN [78] frameworks. Table 2.7 shows the quantitative comparison results in terms of class-wise and overall performance measures. We also present the area under the ROC curve (AUC), average F1-score (F1), and kappa values, along with the average test run time (ms) for the proposed and existing methods. As can be seen in Table 2.7, the proposed method outperforms the existing DL-based methods. Specifically, the 3L-CNN method [72] with the subset of 12-lead ECGs may not be enough to diagnose various localized MIs reliably [21]. It results in poor class-wise and overall measures, as seen in Table 2.7. The MFB-CNN [77] and the MFB-CBRNN [78] exploit 12-lead ECGs for the diagnosis. Although these models show promising

Table 2.7: Performance evaluation of the proposed MLDA-RNN and the existing DL-based methods on the current five-class ECG dataset (5-folds).

Method	**Class**	**Se**(%)	**Sp**(%)	**Acc**(%)	**OS/OSp/OA** (%)	**AUC/F1**(%)/**Kappa**
3L-CNN [72]	EMI	98.61	98.82	98.77		
	AMI	**86.25**	81.64	82.32	77.07	0.84
	CMI	**41.63**	96.38	82.50	93.98	75.14
	non-MI	86.87	97.77	96.31	75.14	0.68
	HC	72.01	95.28	90.35		
MFB-CNN [77]	EMI	92.24	96.29	95.26		
	AMI	**62.32**	93.37	88.77	82.25	0.89
	CMI	**76.07**	92.86	88.60	95.72	83.13
	non-MI	89.23	99.51	98.14	83.13	0.78
	HC	91.39	96.57	95.47		
MFB-CBRNN [78]	EMI	95.84	97.65	97.19		
	AMI	**70.85**	98.72	94.60	**90.01**	**0.94**
	CMI	**89.90**	95.44	94.03	97.66	**90.98**
	non-MI	96.06	99.63	99.15	**90.98**	**0.88**
	HC	97.35	96.88	96.98		
Proposed	EMI	98.34	99.43	99.15		
	AMI	**94.78**	99.62	98.91	**97.60**	**0.98**
	CMI	**97.64**	98.87	98.56	99.43	**97.65**
	non-MI	97.90	99.71	99.47	**97.79**	**0.97**
	HC	99.33	99.51	99.47		

performance for MI vs. HC classification (Table 2.6), they generate poor classification performance for the current five-class dataset with OA of 83.13% and 90.98%, respectively (Table 2.7). These results suggest that the deep CNN-based models may not effectively capture the temporal dependencies of subtle pathological ECG variabilities associated with the MI severity stages. Compared to the state-of-the-art MFB-CBRNN [78], the proposed MLDA-RNN shows a significant improvement of nearly 7% in OA. The model also exhibits improved AUC, F1, and kappa values of 0.98, 97.65%, and 0.97, respectively.

Figure 2.9 compares the OA and the average test run-time for the proposed and existing methods [72, 77, 78] on the current five-class ECG dataset. The measurements for the run times are taken from the system with 16 GB RAM and Quadro K600 graphics processor. As can be seen, the existing CNN-based methods provide faster diagnosis predictions than the proposed MLDA-RNN approach. However, the low OA of existing methods limits their applicability in real-world clinical applications. Although the sequential computations of RNNs in the proposed method contribute to the slight increase in the average run-time of 8.76 ms, they aid in the effective representation of the temporal ECG data and provide improved OA of 97.79%. Considering the order of the input duration in seconds, 8.76 ms run time is not significantly high to delay the emergent interventions [142]. The acceptable test run time and the impressive OA of the proposed method make it best suitable for the preliminary screening of MI severity stages in hospitals/clinics.

From the comparison results in Table 2.6, it can be observed that the MLDA-RNN has impressive

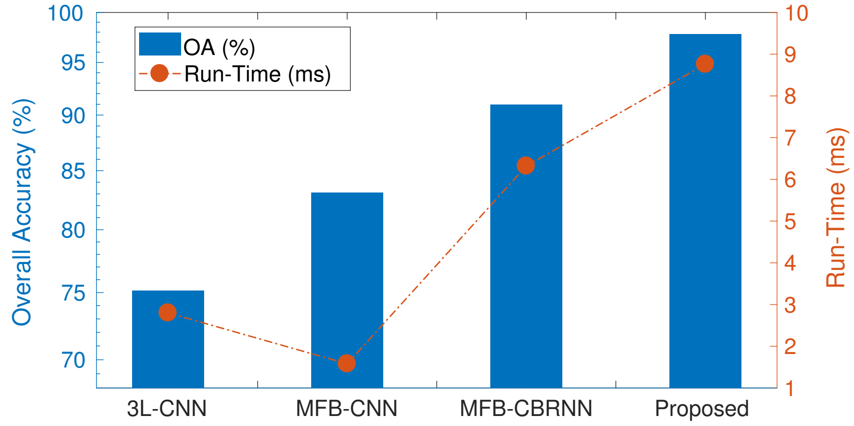


Figure 2.9: Comparison of overall accuracy and the test run-time for the proposed and existing methods.

advantages over other algorithms. (i) We have addressed a challenging problem of MI severity stages classification, where the discriminative changes in the ECG features are subtle. The reliable detection of these stages can help improve disease management with appropriate treatment strategies. (ii) The proposed MLDA-RNN employed RNNs and intra- and inter-lead attention modules to capture the temporal dependencies and fuse diagnostically relevant within- and across-lead ECG information effectively for improved MI staging. (iii) Along with improved diagnostic accuracy, the MLDA-RNN also shows promising model interpretability results through the analysis of intra- and inter-lead attention weights, which are lacking in existing DL models. The analysis of attention weights reveals that the model focuses on the informative ECG leads and features having the pathological manifestations to make the diagnosis decisions. This implies that the model has learned domain-specific clinical knowledge from the 12-lead ECG signals and can provide reliable diagnostic predictions. (iv) Finally, we considered other disease patients through the non-MI class, which reduces over and under treatment due to their misclassification into MI and HCs as in the existing binary classification approaches (MI vs. HC).

Although including other disease patients in the non-MI class improves the reliability of the proposed automated MI staging system, the presence of a high proportion of mimicking MI (M-MI) patients in the non-MI class can affect the performance of MI staging [120]. As discussed before, non-MI patients, such as BBBs, HYP, and myocarditis, show MI-like ECG features across the 12-leads and are often referred to as M-MI diseases. This similarity of ECG features makes reliable MI staging challenging in the presence of such non-MI patients. Table 2.8 presents the distribution of the non-MI class patients in the datasets DS01 and DS02. The DS01 was used in previous experiments to evaluate the proposed MLDA-RNN, consisting of only 16 M-MI patients. On the other hand, DS02 is the newly extracted non-MI dataset from PhysioNet repository, consisting of more M-MI patients, i.e., 29. Table 2.9 shows the detection performance

Table 2.8: Comparison of non-MI class patient's distribution in different datasets.

Cardiac disease	Number of patients in dataset DS01†	Number of patients in dataset DS02
Bundle branch blocks	11	14
Ventricular hypertrophy	4	9
Myocarditis	1	6
	Total mimicking MI patients = **16**	Total mimicking MI patients = **29**
Dysrhythmia	14	14
Valvular heart disease	6	6
Heart failure	18	18
	Total other patients = 38	Total other patients = 38

†: This non-MI class dataset is used for the evaluation of the proposed MLDA-RNN method in previous sections.

Table 2.9: Comparison of the proposed MLDA-RNN method's sensitivity for different non-MI class patient's distribution with EMI, AMI, CMI, and HC classes ECG data unchanged.

Non-MI class dataset	Class	Se(%)	OS(%)
DS01	EMI	98.34	97.60
	AMI	94.78	
	CMI	97.64	
	non-MI	97.90	
	HC	99.33	
DS02	EMI	94.75	90.94
	AMI	82.91	
	CMI	93.68	
	non-MI	84.84	
	HC	98.62	

for five classes with non-MI class patients from DS01 and DS02, with the remaining four classes' ECG data unchanged (see Table 2.1). As can be seen, the presence of more M-MI patients in DS02 significantly affects the detection performance of MI severity stages and non-MI class with a poor OS of 90.94%.

In practice, the clinicians often examine the patient's clinical history along the 12-lead ECG evaluation to make reliable disease predictions and severity assessments [21, 152]. Thus, combining the patient's clinical features with the 12-lead ECG features is expected to improve the MI staging in the presence of mimicking MI patients in non-MI class. However, due to the lack of appropriate clinical history annotations for the MI staging ECG datasets, in the following section, we have investigated a simple MI diagnosis task to evaluate the effectiveness of fusing the patient's clinical features with the 12-lead ECG features.

2.2.7 Fusing Patient's Clinical Features with the 12-Lead ECG for Improving MI Diagnosis

As discussed before, MI is a progressive disease that occurs due to the formation of clots in the coronary arteries, known as atherosclerosis. The INTERHEART study in 52 countries across the globe has identified nine prominent patient risk factors that account for the higher risk of developing MI [152]. Specifically, abnormal high blood-lipid levels, current smoking, and hypertension accelerate atherosclerosis with a relative risk of 3.25, 2.87, and 1.91 times respectively [152]. This study also reported that older adults, especially males are more prone to MI. Therefore, for the first time in the literature, we have investigated the effectiveness of fusing the patient's clinical features such as age, gender, high blood lipid levels, smoking, and hypertension with the 12-lead ECG for improving MI diagnosis in the presence of M-MI patients.

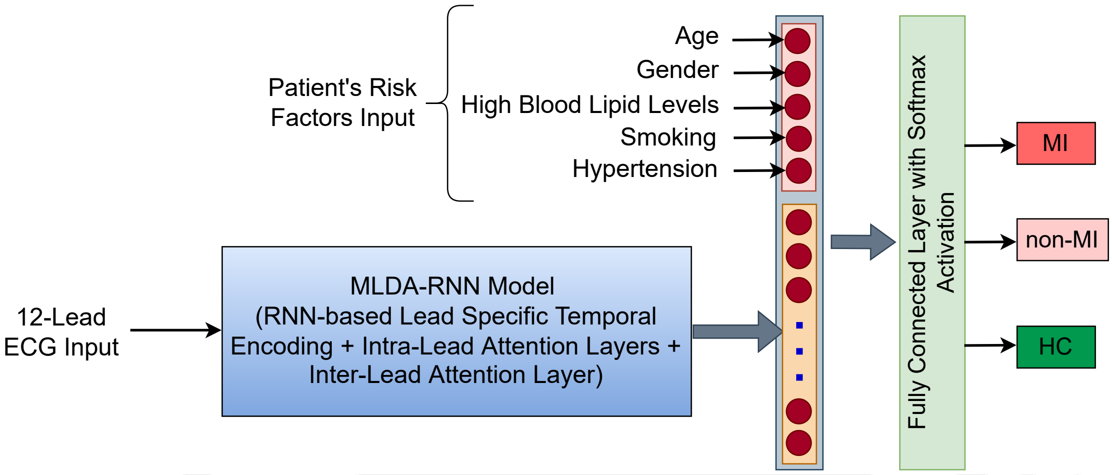


Figure 2.10: The proposed MLDA-RNN-CF architecture for fusing the patient's risk factors with the 12-lead ECG features to classify MI, non-MI, and HC subjects.

The block diagram of the proposed fusion approach, i.e., MLDA-RNN-CF, is shown in Figure 2.10. As can be seen, it considers two types of inputs, i.e., 12-lead ECG and patient's clinical features. The MLDA-RNN model with RNN-based lead-specific temporal encoding followed by intra- and inter-lead attention layers (see Figure 2.2 (a)) is used to extract features from the 12-lead ECG. Its architecture has already been discussed in the previous sections. The 32-dimensional output feature representation from the inter-lead attention layer, i.e., the ALAR vector, is concatenated with the five patient's clinical features input and fed to the output FC+Softmax layer to generate the probability distribution over the MI, non-MI, and HC classes. Generally, the clinical features are available in the form of categorical variables and are converted into numerical in the following manner; gender (1 for male and 0 for female), smoking (1 for smoker and 0 for non-smoker), hypertension (1 for hypertension patients and 0 for non- hypertension patients) and blood lipid level (1 for patients with high blood lipid levels and 0 for low blood lipid levels).

Clinical database and network parameters: The proposed MLDA-RNN-CF model is evaluated on the newly extracted three-class (MI, non-MI, and HC) ECG dataset from the PTB diagnostic database. From this database, we chose individuals with the 12-lead ECG and patient's clinical features annotations for the analysis, and the patients with only ECG data are excluded. The extracted data consists of 124 MI patients,

Table 2.10: Performance comparison of the proposed MLDA-RNN (12-lead ECG input) and the MLDA-RNN-CF (12-lead ECG and patient's clinical features as inputs) methods for the three-class classification task (5 -folds).

Method	**Class**	**Se**(%)	**Sp**(%)	**Acc**(%)	**OS/OSp/OA**(%)	**AUC/Kappa**
	MI	**94.1**$\pm$3.2	95.9$\pm$1.3	95.7$\pm$2.1	**90.6**$\pm$3.2	
MLDA-RNN	non-MI	**80.1**$\pm$5.7	98.9$\pm$0.9	97.1$\pm$1.9	95.6$\pm$1.7	0.953$\pm$0.011
(12-lead ECG)	HC	97.8$\pm$0.9	92.2$\pm$2.5	94.0$\pm$0.9	93.8$\pm$2.9	0.889$\pm$0.027
	MI	**97.6**$\pm$1.1	99.2$\pm$0.6	98.3$\pm$0.5	**98.1**$\pm$0.4	
MLDA-RNN-CF	non-MI	**97.4**$\pm$2.1	98.9$\pm$1.0	99.0$\pm$0.3	99.1$\pm$0.3	0.987$\pm$0.004
(12-lead ECG + clinical features)	HC	99.5$\pm$0.4	99.0$\pm$0.3	99.2$\pm$0.3	98.3$\pm$0.6	0.969$\pm$0.014

49 HC, and 41 non-MI patients (26 M-MI patients and 15 other cardiac disease patients). The dominant portion of the non-MI class consists of M-MI patients, making the reliable MI diagnosis more challenging. Similar to the previous study, we chose the 4 s 12-lead ECG segments for the analysis. A total of 8029 MI, 4427 HC, and 1532 non-MI 12-lead ECG segments with their respective clinical features are used to evaluate the proposed MLDA-RNN-CF model in a patient-independent manner with 5-fold CV. The network is developed using Keras with a TensorFlow backend, and the cost-sensitive multi-class loss is calculated using Eq. 2.14. The network hyperparameters, the training settings, and the evaluation metrics used for MLDA-RNN-CF are similar to the previously discussed MI staging model.

Results: Table 2.10 compares the three-class (MI, non-MI, and HC) classification performance of the MLDA-RNN and the MLDA-NN-CF models in terms of the class-wise and overall evaluation metrics. As can be seen, the model with only 12-lead ECG input, i.e., the MLDA-RNN, shows poor Se for the non-MI and MI classes with an OS of 90.6%. However, the model with 12-lead ECG and patient's clinical features as input, i.e., the MLDA-RNN-CF, significantly improves the Se of non-MI and MI classes with an OS of 98.1%. These results suggest that with the 12-lead ECG features from MLDA-RNN, it is difficult to discriminate between MI and non-MI classes (with more mimicking MI patients) due to the ECG features similarity among the classes. However, the fusion of the patient's clinical features to the 12-lead ECG features introduces complementary information to the model, which aids in better discrimination between the MI and non-MI classes, thereby achieving improved classification performance.

2.3 Summary

The automated severity staging of MI can help assess the disease progression and assist cardiologists in improving the disease management by timely initiating appropriate treatment strategies. To this end, in this chapter, we presented a unique MLDA-RNN architecture for automated classification of MI severity stages such as early MI, acute MI, and chronic MI from the 12-lead ECG signals. The main contribution presented in this chapter is the exploration of RNNs and the self-attention modules for effective feature extraction from the 12-lead ECG to improve MI staging. Specifically, the MLDA-RNN model systematically processes the 12-lead ECG using the RNN-based temporal encoding layers followed by intra- and inter-lead attention layers. The RNN layers effectively model the temporal dependencies of the subtle pathological ECG manifestations associated with the progressive MI severity stages. The intra- and inter-lead attention layers automatically emphasize the clinically relevant within- and across-lead diagnostic information of the 12-lead ECG, respectively, for improved classification. This systematic feature fusion also helps alleviate the diagnostic redundancy of the 12-lead ECG to learn more robust feature representations than the early ECG-lead fusion model, i.e., MLDA-RNN-E. Along with the improved performance, the notable advantage of the MLDA-RNN method is the diagnostic transparency provided by the attention weights. The visualization of intra- and inter-lead attention weights demonstrates the correlation of weights with the diagnostic relevance of focused ECG features and leads, which often aligns with medical knowledge and correlates with the cardiologist's way of 12-lead ECG examination.

In this chapter, we also presented the effectiveness of fusing the patient's clinical features with the 12-lead ECG to improve MI and non-MI diagnosis in the presence of mimicking MI patients. The preliminary findings suggest that incorporating the patient's clinical history and the 12-lead ECG into clinical evaluation has the potential to improve the diagnosis. Thus, the proposed MI diagnosis method (MLDA-RNN-CF) can be reliably used to pre-screen MI patients at clinics/hospitals for referring them to a cardiologist. Subsequently, the proposed MI staging approach (MLDA-RNN) can assess the MI progression and assist cardiologists in timely initiating suitable emergent interventions.

3

Congestive Heart Failure Diagnosis from Single-Lead ECG using Attention-based Deep Residual RNN

Contents

Congestive heart failure (CHF) is a complex clinical syndrome in which the heart fails to meet the body's circulatory demands. It can result from any cardiac disorder that impairs ventricular filling or pumping of the blood, including MI, HYP, BBBs, valvular heart disease, and cardiac arrhythmia. CHF is a serious medical condition associated with higher incidence and mortality rates and reduced health-related quality of life (QoL) [33]. Therefore, early and accurate detection of CHF is critical for initiating appropriate treatment plans to improve the heart pumping efficiency and health-related QoL. Clinicians often prefer the non-invasive ECG test for the preliminary diagnosis of CHF in hospitals and large-scale screening programs. Because the etiology of the CHF is diverse, there are no definite ECG guidelines for its diagnosis. However, the ANS damage during CHF causes reduced HRV derived from the RR-intervals of the ECG. Most existing automated methods [81–85] employ HRV indices to characterize CHF patients. Generally, HRV feature-based approaches require long-term ECG of more than 24 h duration or at least 5-min/30-min for the reliable assessment of CHF, which is often difficult to acquire using portable mobile ECG devices. Notably, the HRV features are only effective in identifying severe CHF patients (NYHA III and IV) with acute ANS damage and thus are unreliable for diagnosing patients at the early stages (NYHA I and II).

As discussed in the introduction chapter, apart from HRV, specific pathological ECG patterns, including ST-segment elevation/depression, prolonged QRS- and QT-duration, deep S-wave, and T-wave abnormalities, are often associated with the CHF. To automatically detect the CHF from the ECG signals, the method must implicitly recognize these complex pathological ECG features and discern temporal relationships between them. In this direction, few recent methods employed deep CNN-based feature representation models [49, 88, 89], which may not adequately capture complex ECG features and their temporal dependencies associated with the CHF syndrome. Moreover, these deeper networks are hard to optimize because of the vanishing gradient problem. The existing DL methods lack ”diagnostic transparency,” which limits the applicability of such ”blackbox” methods in real-world clinical applications. Although Porumb *et al.* [49] have presented model transparency using the Grad-CAM technique, it only helps visualize the salient ECG features that led to the CHF detection during the testing stage. It would be interesting if we could constrict the model to learn to focus on the diagnostically relevant ECG features during the training stage. Recently, incorporating residual connections between the layers has been shown to ease the training of deeper networks and aid in learning complex representations for improving image and speech recognition [144, 153], and arrhythmia classification [15, 57] performance. Considering the efficacy of RNNs and attention networks for modeling sequential 12-lead ECG from the previous chapter and residual connections from [15], we presume that designing a hybrid network using these components can learn the complex ECG features related to CHF and provide improved diagnosis and model transparency results.

This chapter proposes an attention-based deep residual RNN (A-DRRNet) architecture to diagnose CHF from single-lead ECG beats. The model employs multi-layered RNNs consisting of recurrent hidden states to effectively encode temporal dependencies (P-QRS-T) of complex pathological ECG features. In addition, residual connections have been introduced between the RNNs layers to improve the network's ability to learn complex features hierarchically by maintaining the proper gradients at the initial layers. Finally, incorporating the self-attention module further improves the feature learning pipeline by focusing on clinically relevant ECG features for efficient CHF detection. The rest of the chapter is organized as follows. Section 3.1 and 3.2 present the proposed framework and the experimental results of the A-DRRNet method, respectively. The summary of the chapter is presented in section 3.3.

3.1 Attention based Deep Residual RNN Approach for CHF diagnosis

The block diagram of the proposed A-DRRNet framework is presented in Figure 3.1. The method mainly consists of three processing modules: the deep residual RNN-based temporal encoding module, the attention module, and the classification module. First, the incoming ECG beat data is fed to the temporal encoding module consisting of multilayered residual RNNs that effectively model the temporal dynamics of sequential ECG data and extract powerful high-level residual representation vectors. The attention module generates weights for the residual vectors based on their diagnostic relevance, aggregating them through weighted summation to form a beat-attentive representation vector. The beat-attentive vector is finally applied to the classification module to generate a diagnosis decision for the input ECG beat. The following sub-sections provide the details of each of the modules.

3.1.1 Deep Residual RNN-based Temporal Encoding

As discussed before, the sequential activation of the heart's electrical conduction system results in the ECG heartbeat; thus, temporal dependencies naturally exist in this waveform. RNNs are a class of artificial neural networks (ANN) widely used for modeling the sequential data [147]. It processes the sequential inputs by having a recurrent hidden state whose activation at the present time-step depends on the previous time-step [147]. In this study, we choose the gated recurrent unit (GRU) as the basic RNN unit to effectively model the time dependencies of the ECG beat data associated with CHF syndrome.

3.1.1.1 Low-level Feature Extraction

Let the incoming ECG heartbeat sequence be denoted as $\mathbf{x} = [x_1, x_2, ..., x_t, ..., x_T]$. Here, x_t represents heartbeat information at the time-step t, $t \in \{1, 2, ..., T\}$ with T represents the sequence length. The GRU

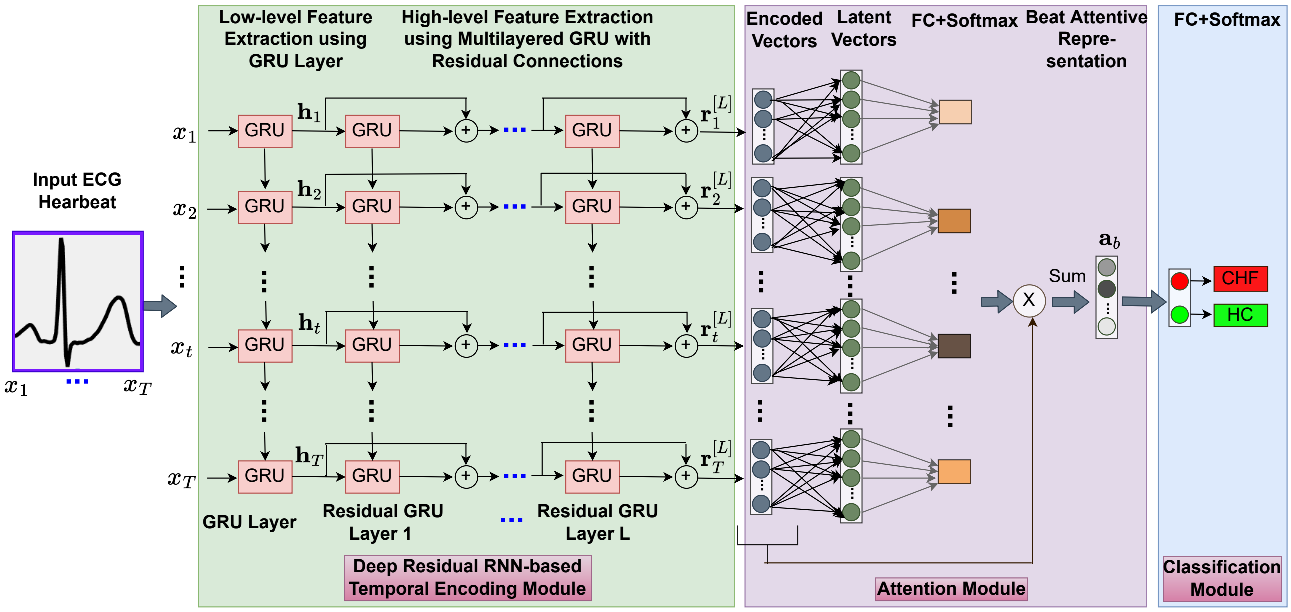


Figure 3.1: Block diagram of the proposed A-DRRNet architecture for detecting CHF form single-lead ECG beat.

cell consists of two important gates known as reset gate ($\mathbf{g}$) and update gate ($\mathbf{z}$) for regulating the flow of information (see Figure 3.2 (a)). The computations of the GRU cell at time-step t can be given as follows:

$$\mathbf{g}_t = \sigma(\mathbf{w}_{ri}x_t + \mathbf{W}_{rh}\mathbf{h}_{t-1} + \mathbf{b}_r) \tag{3.1}$$

$$\mathbf{z}_t = \sigma(\mathbf{w}_{zi}x_t + \mathbf{W}_{zh}\mathbf{h}_{t-1} + \mathbf{b}_z) \tag{3.2}$$

$$\widetilde{\mathbf{h}}_t = tanh(\mathbf{w}_{ci}x_t + \mathbf{W}_{cr}(\mathbf{g}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_c) \tag{3.3}$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \widetilde{\mathbf{h}}_t \tag{3.4}$$

where $\mathbf{h}_t$ and $\mathbf{h}_{t-1}$ represents the hidden state at time step t and $t-1$, respectively and $\widetilde{\mathbf{h}}_t$ represnts candidate hidden state. $\sigma(\cdot)$ is the sigmoid activation function, $tanh$ is the hyperbolic tangent activation function and $\odot$ represents element-wise multiplication. $\mathbf{W}_{rh}$, $\mathbf{W}_{zh}$ and $\mathbf{W}_{cr}$ are recurrent weight matrices, $\mathbf{w}_{ri}$, $\mathbf{w}_{zi}$ and $\mathbf{w}_{ci}$ are weight vectors (matrices in residual GRU layers) and $\mathbf{b}_r$, $\mathbf{b}_z$ and $\mathbf{b}_c$ are bias vectors.

For the input ECG beat sequence $\mathbf{x} \in \mathbb{R}^T$ its encoded hidden vectors from the first GRU layer are $\mathbf{h}_t \in \mathbb{R}^{d_h}$, $t \in \{1, 2, ..., T\}$, where d_h represents the encoding size or GRU memory. The encoded vectors at the first GRU layer are limited in the number of non-linearities applied to the input ECG, thus learning low-level feature representations from the sequential ECG data. These low-level features are fed to the multilayered GRU with residual connections to hierarchically extract the temporal dependencies of high-level or complex pathological ECG characteristics.

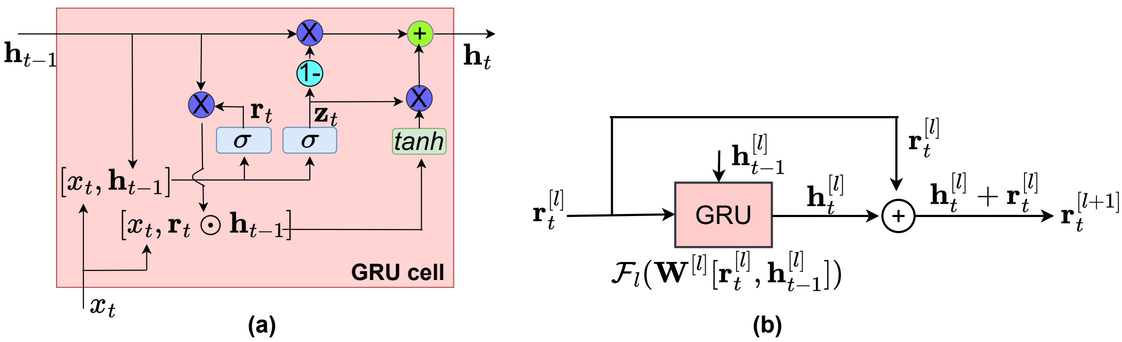


Figure 3.2: (a) GRU cell architecture. (b) A typical GRU cell with residual connection.

3.1.1.2 High-level Feature Extraction using Multilayered GRU Architecture with Residual Connections

In order to learn complex features, a common way used in neural networks is to stack more layers that can extract high-level feature representations hierarchically. However, the deeper networks are hard to optimize because of the vanishing gradient problem, i.e., as the gradient is back-propagated to earlier layers, repeated multiplication makes the gradient significantly small and results in slow convergence and performance degradation of the model [153]. Recently Microsoft Research Team has developed a residual learning framework to improve the ease of training, and representation ability of the deeper networks [153]. Residual leaning introduces an identity skip/residual connections between adjacent layers. These connections alleviate the vanishing gradient problem by propagating sufficient gradients to the initial layers so that these layers can also learn as fast as the final layers.

Inspired by the idea of residual learning, in the proposed network, we incorporate residual connections from one GRU layer to the next (see Figure 3.1) to extract effective high-level representations. Figure 3.2 (b) shows a typical GRU cell with residual connection at time step t and residual layer l, $l \in \{1, 2, .., L\}$, where L represents the total number of residual GRU layers. Its formulation is given as follows:

$$\mathbf{h}_t^{[l]} = \mathcal{F}_l(\mathbf{W}^{[l]}[\mathbf{r}_t^{[l]}, \mathbf{h}_{t-1}^{[l]}]) \tag{3.5}$$

$$\mathbf{r}_t^{[l+1]} = \mathbf{h}_t^{[l]} + \mathbf{r}_t^{[l]} \tag{3.6}$$

$$\mathbf{h}_t^{[l+1]} = \mathcal{F}_{l+1}(\mathbf{W}^{[l+1]}[\mathbf{r}_t^{[l+1]}, \mathbf{h}_{t-1}^{[l+1]}]) \tag{3.7}$$

where $\mathcal{F}_l$ represents the GRU cell function (Eq. 3.1-3.4) with corresponding weight matrix $\mathbf{W}^{[l]}$ and it generates the current hidden vector $\mathbf{h}_t^{[l]}$ (Eq. 3.5) as a function of two inputs, i.e., $\mathbf{r}_t^{[l]}$: the current residual input and $\mathbf{h}_{t-1}^{[l]}$: the previous time step hidden vector. Next, using the skip connection, the input $\mathbf{r}_t^{[l]}$ vector is element wise added to the current hidden state $\mathbf{h}_t^{[l]}$ to obtain the output residual vector $\mathbf{r}_t^{[l+1]}$ (Eq. 3.6). This $\mathbf{r}_t^{[l+1]}$ vector is fed as a input to the next layer's $(l + 1)$ residual GRU cell. Each residual layer's hidden vectors and residual outputs are generated in an iterative manner following the Eq. 3.7. Following this formulation, the deep residual RNN model can be designed by stacking multiple such GRUs with residual connection blocks one after the other, with the output of the previous block forming the next's input (see Figure 3.2 (b)). The output of the last layer's residual GRU block can be computed as:

$$\mathbf{r}_t^{[L]} = \mathbf{h}_t^{[L-1]} + \mathbf{r}_t^{[L-1]} \tag{3.8}$$

The output residual vectors from the last residual GRU layer (L) are next fed to the attention module.

3.1.2 Attention Module

To detect CHF patients, in practice, cardiologists focus on the diagnostically informative ECG characteristics and examine their variations relative to the HC subjects [79]. In the previous chapter, the attention networks have shown promising results for MI severity staging by emphasizing the salient discriminative features of the 12-lead ECG during the classification process. Therefore, in this study also, we have employed a self-attention module on top of the deep residual RNN architecture (see Figure 3.1) to improve discriminative feature learning and model transparency analysis. This module generates the weights of importance for each residual representation vectors ($\mathbf{r}_t^{[L]}$) based on their clinical relevance in the diagnosis and fuse them to form a beat attentive representation (BAR) vector $\mathbf{a}_b \in \mathbb{R}^{d_h}$. For this, first, each residual representation $\mathbf{r}_t^{[L]}$ is non-linearly transformed into a latent-space using a fully connected (FC) layer as:

$$\mathbf{u}_t = tanh(\mathbf{W_{att}}\mathbf{r}_t^{[L]} + \mathbf{b_{att}}) \tag{3.9}$$

where $\mathbf{W_{att}} \in \mathbb{R}^{d_a \times d_h}$ is weight matrix and $\mathbf{b_{att}} \in \mathbb{R}^{d_a}$ is bias vector of the FC layer with latent size d_a, which maps $\mathbf{r}_t^{[L]}$ to a latent representation $\mathbf{u}_t \in \mathbb{R}^{d_a}$. Subsequently, the Softmax activation is applied to the $\mathbf{u}_t$, which generates the weight of importance for each $\mathbf{r}_t^{[L]}$ computed as:

$$\alpha_t = \frac{exp(\mathbf{u}_a^{\mathsf{T}}\mathbf{u}_t + b_a)}{\sum_{t=1}^{T} exp(\mathbf{u}_a^{\mathsf{T}}\mathbf{u}_t + b_a)} \tag{3.10}$$

where $\mathbf{u}_a \in \mathbb{R}^{d_a}$ and $b_a \in \mathbb{R}^1$ are the trainable parameters of the network. The dot-product similarity between $\mathbf{u}_t$ and $\mathbf{u}_a$ generates attention weight α_t, $t \in \{1, 2, ..., T\}$ at each time-step. The Softmax function ensures that all the computed weights sum to 1, and this weight of importance values will be relatively high for the residual vectors, which are more diagnostically relevant for classification than others. Finally, a weighted sum of all the temporal residual vectors is calculated to form a high-level BAR vector given as:

$$\mathbf{a}_b = \sum_{t=1}^{T} \alpha_t \mathbf{r}_t^{[L]} \tag{3.11}$$

where $\mathbf{a}_b \in \mathbb{R}^{d_h}$ denotes the BAR vector obtained by emphasizing the diagnostically discriminative residual vectors of the ECG sequence to improve CHF diagnosis (see Figure 3.1).

3.1.3 Classification Module

In the classification stage, the $\mathbf{a}_b$ vector is fed to the FC+Softmax output layer to obtain the probability distribution of outputs that the input ECG heartbeat $(\mathbf{x})$ belongs to each class computed as:

$$p(c|\mathbf{x}) = Softmax(\mathbf{W}_o\mathbf{a}_b + \mathbf{b}_o) \tag{3.12}$$

where $p(c|\mathbf{x})$ is the probability of input $\mathbf{x}$ belongs to class c, $c \in \{1, 2, ..., C\}$ with C denote number of classes under study, i.e., two (CHF and HC), and $\mathbf{W}_o \in \mathbb{R}^{C\times d_h}$, $\mathbf{b}_o \in \mathbb{R}^C$ are weight matrix and bias vector of the output layer, respectively. The final predicted output class label for the input beat $\mathbf{x}$ is obtained as:

$$class(\mathbf{x}) = \underset{c=1,2,\ldots,C}{\arg\max}\, p(c|\mathbf{x}) \tag{3.13}$$

3.1.4 Training Model Parameters

In the training phase, all the model parameters including the GRU layer, the residual GRU layers, the attention module and the output layer are trained using backpropagation by minimizing the binary cross-entropy loss $(\mathcal{L})$ between the conditional probability output and the true labels for a mini-batch training samples $(\mathbf{x}_n, y_n)$, $n \in \{1, 2, ..., N\}$ given as:

$$\mathcal{L} = -\frac{1}{N}\sum_{n=1}^{N}\big(y_n log(p(c|\mathbf{x}_n)) + (1-y_n)log(1-p(c|\mathbf{x}_n))\big) \tag{3.14}$$

where y_n denote the true label of the n^{th} training sample $\mathbf{x}_n$ and N denote the mini-batch size.

3.1.5 Analysis of Deep Residual RNNs Backpropagation Properties

In general, RNNs are deep in temporal order because of their recurrent transition from one hidden state to another. With the depth in time, the proposed deep residual RNN model is also deep along the layer structure. However, to emphasize the effectiveness of residual connections in the gradient analysis, we mainly focus on the gradient backpropagation along the depth of layers. Through recursively updating Eq.

3.6, we can formulate forward information flow between any shallower layer l and deeper layer L as follows:

$$\mathbf{r}_t^{[L]} = \mathbf{r}_t^{[L-1]} + \mathcal{F}_{L-1}(\mathbf{W}^{[L-1]}[\mathbf{r}_t^{[L-1]}, \mathbf{h}_{t-1}^{[L-1]}]) \tag{3.15}$$

$$= \mathbf{r}_t^{[L-2]} + \mathcal{F}_{L-2}(\mathbf{W}^{[L-2]}[\mathbf{r}_t^{[L-2]}, \mathbf{h}_{t-1}^{[L-2]}]) + \mathcal{F}_{L-1}(\mathbf{W}^{[L-1]}[\mathbf{r}_t^{[L-1]}, \mathbf{h}_{t-1}^{[L-1]}]) \tag{3.16}$$

$$= \mathbf{r}_t^{[l]} + \sum_{i=l}^{L-1} \mathcal{F}_i(\mathbf{W}^{[i]}[\mathbf{r}_t^{[i]}, \mathbf{h}_{t-1}^{[i]}]) \tag{3.17}$$

where Eq. 3.17 indicates that the output feature vector $\mathbf{r}_t^{L}$ of any deeper layer L can be represented as the feature vector $\mathbf{r}_t^{[l]}$ of any shallower layer l plus a residual function in the form $\sum_{i=l}^{L-1} \mathcal{F}_i$, indicating that the deep network architecture is in a residual fashion between any layers L and l. Eq. 3.17 has nice backpropagation properties. As the model's loss function is $\mathcal{L}$ (Eq. 3.14), by the chain rule of backpropagation, the gradients at any shallower layer l can be computed as:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[l]}} = \frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[L]}} \frac{\partial \mathbf{r}_t^{[L]}}{\partial \mathbf{r}_t^{[l]}} \tag{3.18}$$

$$= \frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[L]}} \left[\frac{\partial}{\partial \mathbf{r}_t^{[l]}} \left[\mathbf{r}_t^{[l]} + \sum_{i=l}^{L-1} \mathcal{F}_i(\mathbf{W}^{[i]}[\mathbf{r}_t^{[i]}, \mathbf{h}_{t-1}^{[i]}]) \right] \right] \tag{3.19}$$

$$= \frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[L]}} \left[1 + \frac{\partial}{\partial \mathbf{r}_t^{[l]}} \left[\sum_{i=l}^{L-1} \mathcal{F}_i(\mathbf{W}^{[i]}[\mathbf{r}_t^{[i]}, \mathbf{h}_{t-1}^{[i]}]) \right] \right] \tag{3.20}$$

$$= \frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[L]}} + \frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[L]}} \left[\frac{\partial}{\partial \mathbf{r}_t^{[l]}} \left[\sum_{i=l}^{L-1} \mathcal{F}_i(\mathbf{W}^{[i]}[\mathbf{r}_t^{[i]}, \mathbf{h}_{t-1}^{[i]}]) \right] \right] \tag{3.21}$$

where Eq. 3. 21 shows that the gradient $\frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[l]}}$ can be decomposed into two additive terms, i.e., a term of $\frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[L]}}$ that backpropagate gradients information directly without concerning any weight layers, and another term of $\frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[L]}}(\frac{\partial}{\partial \mathbf{r}_t^{[l]}} \sum_{i=l}^{L-1} \mathcal{F}_i)$ that propagates through the weight layers. The first term of $\frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[l]}}$ ensures that gradients are directly propagated back to any shallower layer l. Eq. 3.20 demonstrates that it is unlikely for the gradient $\frac{\partial \mathcal{L}}{\partial \mathbf{r}_t^{[l]}}$ to be cancelled out for a mini-batch data because in general $\frac{\partial}{\partial \mathbf{r}_t^{[l]}} \sum_{i=l}^{L-1} \mathcal{F}_i$ cannot be always -1 for all samples in a mini-batch. This implies that the gradient at any shallower layer l does not vanish even when the intermediate layers weights are arbitrarily small. This impressive backpropagation property of the residual connections allows us to train deeper architectures that exhibits better representation ability.

3.2 Experimental Results and Discussion

This section presents the experimental setup and the experimental results of the proposed A-DRRNet method. This section also presents the analysis of the proposed A-DRRNet model, followed by the performance comparison with the existing methods.

3.2.1 Experimental Setup

3.2.1.1 *Clinical ECG Database*

The CHF's severity can be categorized into four levels (I-IV), and existing approaches focus only on identifying severe CHF patients (III and IV) from HCs. In this study, three ECG datasets [148], namely the Beth Israel Deaconess Medical Centre Congestive Heart Failure Database (BIDMC-CHF), the PTB Diagnostic Database (PTBDB), and the MIT-BIH Normal Sinus Rhythm Database (NSRDB), are used for the analysis. As shown in Table 3.1, a total of 73 subjects (CHF: 35; HC:38) ECG recordings are extracted and used for the model evaluation. Specifically, the CHF data in the current study consists of 35 patients with severity levels ranging from I to IV. Achieving reliable classification results on the diverse ECG datasets is challenging considering the more patients and varying severity levels. The long-term ECG recordings of the datasets are acquired using ambulatory ECG recorders. These recorders provide practical and real-world ECG signals. Also, the datasets are acquired from subjects having different gender (40 male and 33 female subjects) and age ranges (20-80 years). In all databases, we choose ECG lead-I for data extraction and consider ECG beat-level diagnosis. Table 3.1 summarizes the details of the extracted data.

Table 3.1: Details of ECG signals extracted from various publicly available databases.

Database	Sampling rate	Duration	Class	NYHA-level	#Subjects	#Males (age)	#Females (age)	#Extracted beats
BIDMC-CHF	250 Hz	~20 h	CHF	III-IV	15	11 (22-71 years)	4 (54-63 years)	202,330
PTBDB	1000 Hz	30s - 2 min		I-III	20	11 (39-82 years)	9 (32-80 years)	
MIT-BIH NSRDB	128 Hz	~24 h	HC	N/A	18	5 (26-45 years)	13 (20-50 years)	217,348
PTBDB	1000 Hz	30s - 2 min		N/A	20	13 (25-39 years)	7 (28-41 years)	

#: Number of; N/A: Not applicable.

3.2.1.2 *Data Preprocessing*

To remove the baseline artifact from the ECG signals, we have employed a second-order Butterworth low pass filter with a cut-off frequency of 0.5 Hz. Implementation of this noise removal technique consists of two steps; first, the incoming original noisy ECG signal is passed through the LPF filter, which filters out the BW noise; later, the filtered BW noise is subtracted from the original noisy ECG to obtain the noise-free signal. **Beat extraction:** To reduce the computational burden, all the ECG signals are down-sampled to 128 Hz. ECG recordings from the databases contain beat annotation files, which were used to extract individual beats. We chose a 235 ms window before the R-peak and a window of 390 ms after the R-peak to extract the heartbeats [49]. **Normalization:** The ECG recordings amplitudes vary among different subjects and databases. However, neural networks tend to converge faster and produce better generalization results when all the inputs have similar distributions. Therefore, each ECG beat is z-score normalized to obtain zero mean and unit variance. As the BIDMC and NSRDB contain long-term ECG recordings (minimum 20h), the number of heartbeats extracted from each subject was close to 70,000. The temporally close beats contain similar patterns, so we randomly selected a single heartbeat every 8 s of the recording. The details of the number of beats used in the study are shown in Table 3.1.

3.2.1.3 Evaluation Strategies

We employed two diagnosis strategies: i) beat-level and ii) 5 min ECG excerpts-level. The proposed approach performs beat-level detection and is extended to excerpts-level classification through the "majority voting technique" [49]. Specifically, let N_1 and N_2 be the number of beats classified as HC and CHF within 5min of the ECG excerpt, respectively. The final diagnosis decision for the given 5 min is calculated as $max(N_1, N_2)$. We use the standard binary classification evaluation measures [80], including Se, Sp, Acc, and AUC, to evaluate the performance of the proposed method.

3.2.1.4 Data Splitting Procedure

The extracted dataset is randomly divided into three subsets, i.e., training, validation and testing, in the proportion of about 50$\%$, 25$\%$ and 25$\%$ of the total data, respectively. As the ECG beats of an individual subject show similarity, we have employed a patient-independent scheme that ensures no overlap of the same subject's ECG heartbeats while dividing three subsets. To alleviate over-fitting and maintain balance bias and variance for the model, we repeated the random data dividing process for 10 runs. The proposed model trained and evaluated for each run and the mean-variance of the results across the 10 runs are reported as the model's final performance.

3.2.1.5 Hyper-Parameter Selection

To achieve better generalization results for DL-based networks, the proper selection of hyper-parameters is crucial. The proposed architecture has three important hyper-parameters, i.e., the number of residual GRU layers (L), the GRU encoding size (d_h) and the latent-space dimension (d_a). These parameters are optimized by using a grid search procedure. For this, we varied L in the range 1 to 6, d_h values in the range 4 to 64 and d_a values in the range 4 to 128. As can be seen from Figure 3.3 (a) and (b), for $L = 4$, $d_h = 8$ and $d_a = 16$, the proposed method achieves consistently good results with minimum number of trainable parameters. We experimentally found that the first GRU layer without residual connections provides better results by capturing low-level representations (see Figure 3.1). This also helps to avoid additional parameters to match dimensions for addition in residual blocks.

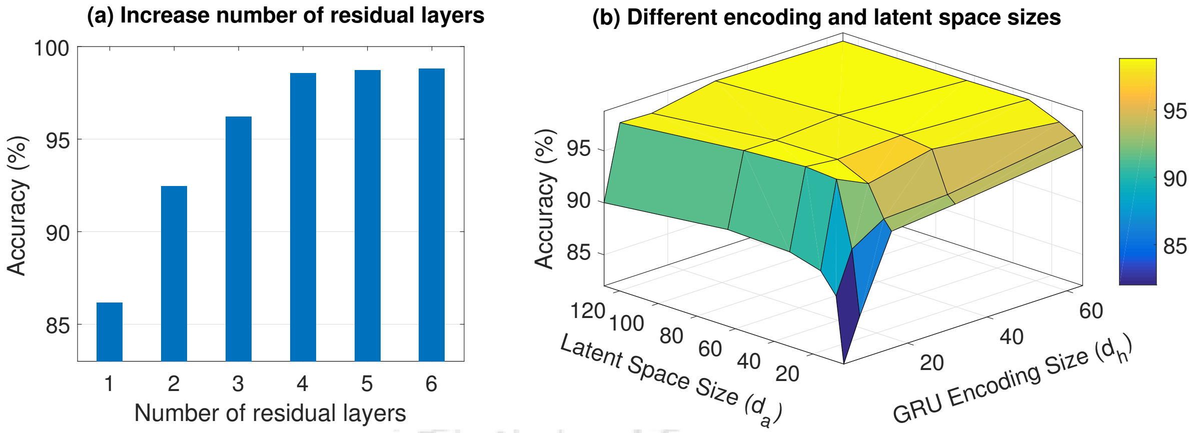


Figure 3.3: Proposed A-DRRNet architecture optimization.

The complete network is implemented in Keras using the TensorFlow backend. The loss function is calculated using Eq. 3.14, and the model is trained using backpropagation through time. We employed Adam optimizer as the optimization method with a learning rate of 10^{-3} (set experimentally). The batch size is set to 32, and the model is trained for 5 epochs. An early stopping criterion using the validation dataset is employed to stop training when the AUC does not improve over 20 iterations. A drop-out layer with a drop-rate of 0.25 is applied to the output layer to reduce over-fitting.

Table 3.2: Beat-level mean confusion matrix for the test data (10 runs).

	True CHF	**True HC**	**Average Error** (%)
Classified as CHF	68278	1173	-
Classified as HC	624	58965	-
Class-wise error	∼1%	∼2%	∼1.50%

Table 3.3: Performance in terms of mean±standard deviation of the metrics on the test data (10 runs).

Diagnosis	**Se**(%)	**Sp**(%)	**Acc**(%)	**AUC**
Beat-level	99.01 ± 0.58	97.89 ± 1.54	98.57 ± 0.99	0.98 ± 0.01
5min excerpt-level	99.82 ± 0.01	98.90 ± 1.05	99.37 ± 0.04	0.99 ± 0.007

3.2.2 Experimental Results

Table 3.2 and 3.3 show the mean confusion matrix and the average beat-level performance of the proposed method on the test data over ten runs. It can be seen from Table 3.2 that for the beat-level diagnosis, the proposed method achieves an impressive Se of 99.01%, Sp of 97.89%, Acc of 98.57% and AUC of 0.98. It can be seen from Table 3.3 that the number of false positives (FP) is around only 2% of the mean number of HC heartbeats, while the false negatives (FN) are around only 1% of the mean number of CHF heartbeats.

A closer look at the misclassified beats in one of the random runs reveals that most of the misclassified CHF beats are from the patient id:chf08. Similarly, the majority of the misclassified HC beats are from the subject id:16795. It should be noted that these recordings contain long-term (∼20 h) ECG signals, and the misclassified beats represent only 1%-3% of the total heartbeats. Therefore, such beat-level misclassification can be remedied using a majority voting scheme for the record-level diagnosis as more than 90% beats are correctly classified in these records. To adapt the proposed beat-level diagnosis method for the continuous patient monitoring in hospital-care or home-healthcare applications, we have employed a majority voting-based diagnosis for heartbeats within 5 min excerpts. As discussed earlier, the class associated with the 5 min ECG excerpt is the majority class of the individual heartbeats in that excerpt. Table 3.3 shows the average classification results for the 5 min excerpts using the majority voting scheme. As can be seen, all the measures are further improved over an individual beat-level diagnosis, suggesting an accurate model for screening CHF patients.

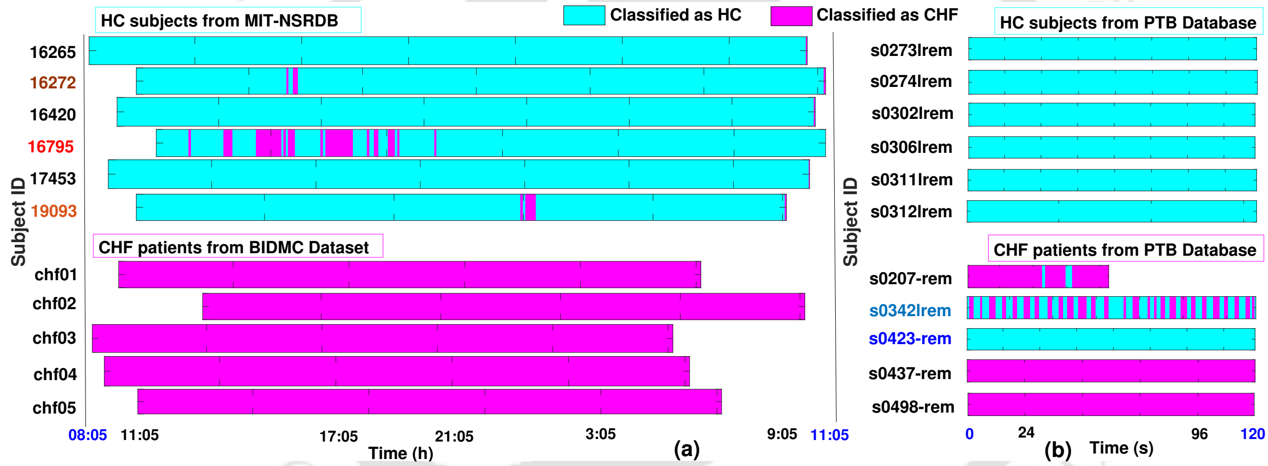


Figure 3.4: Examination of classification results for one of the random runs test data. (a) Presents majority voting-based 5min excerpt-level classification results for the long-duration ECG recordings. (b) Presents beat-level classification results for the PTBDB (time-scale increased for better viewing).

Figure 3.4 presents the visual representation of record-level diagnosis mapping for the test data of one of the random runs. Here teal and magenta color coding represent the classified results as HCs and CHF obtained by majority voting over 5 min segments for the long-term ECG recordings (x-axis of Figure 3.4 (a)) of different subjects (y-axis of Figure 3.4 (a)), respectively. As the PTBD database ECG recordings vary between the 30 s - 2 min, we have shown only beat-wise detection results (Figure 3.4 (b)). As can be seen, most of the ECG records from the test data are correctly classified. Specifically, all the 5 min segments of the CHF patients from the BIDMC database are accurately classified (Figure 3.4 (a)). Also, all the HC beats from the PTBD database were correctly classified (Figure 3.4 (b)). However, HCs from MIT-NSRDB and CHF patients from PTBDB have few misclassified segments/beats. We have plotted

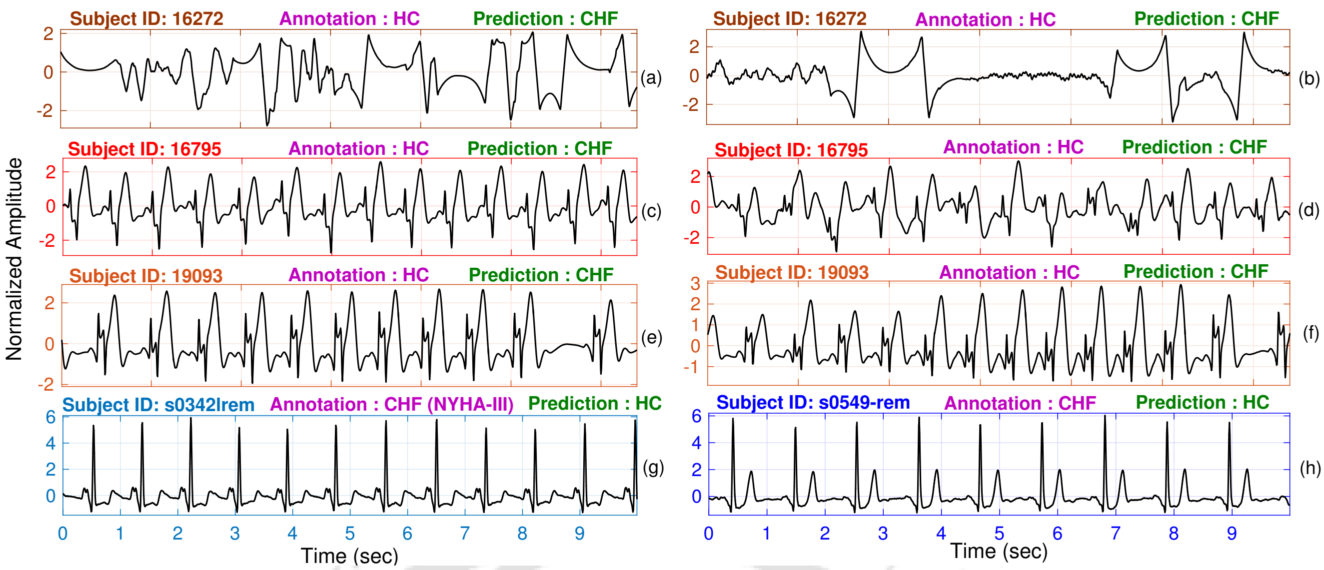


Figure 3.5: Typical examples of the misclassified ECG beats/segments for the test dataset. (a)-(f) Show the ECG segments from subjects IDs 16272, 16795, and 19093 misclassified as CHF. (g)-(h) Show the ECG segments from subjects IDs s0342lrem and s0549-rem misclassified as HCs.

these misclassified ECG segments in Figure 3.5 by highlighting the annotation label and prediction label for a better understanding of the decisions made by the proposed method. As observed, misclassified segments from HC subject id:16272 are severely affected by poor electrode connection noise (Figure 3.5 (a) and (b)) and do not represent HCs ECG morphologies. Also, the misclassified segments from HC subjects id:16795 (Figure 3.5 (c) and (d)) and id:19093 (Figure 3.5 (e) and (f)) contains ST-segment and T-wave abnormalities along with slight movement artifacts. The misclassified segments from CHF patients id:s0342lrem (Figure 3.5 (g)) and id:s0549-rem (Figure 3.5 (f)) contains ECG features similar to the HCs. Therefore, apart from a few misclassified ECG segments containing either artifacts (id:16272 and id:16795) or erroneous annotations (id:16795, id:19093, and id:s0549-rem), the results from Figure 3.4 and Table 3.3 indicate that the proposed beat-level diagnosis achieves nearly accurate results for the CHF analysis. These results also demonstrate the effectiveness of the proposed method for adapting to the continuous patient monitoring applications using long-term ECG recordings.

3.2.3 Effectiveness of the Proposed A-DRRNet Method

In this section, we analyze the effectiveness of each component of the proposed method on the overall performance. For this, we perform an ablation study by evaluating some variants of the proposed method. This section also presents the diagnostic transparency of the proposed approach through the visualization of the learned attention weights corresponding to the ECG features during the diagnosis process.

3.2.3.1 Significance of the Deep Residual RNN Architecture

To examine residual connection's effectiveness, we designed a deep RNN (DRNN) architecture without residual connections named plane DRNN and with residual connections named deep residual RNN. The plane DRNN contains a stack of five GRU layers, whereas the deep residual RNN contains one GRU layer followed by four residual GRU layers (see Figure 3.1). Table 3.4 compares the plane and residual DRNN models. As can be seen, the DRNN model with residual connections performs better than the plane DRNN with a minimum improvement of nearly 5% in Se, 7% in Sp, 6% in Acc, and 6% in AUC. This increase in performance is because the residual connections improve the ease of training and representation ability of the deep RNN architecture by maintaining proper gradients at the initial layers. These results verify the mathematical derivations of the deep residual RNNs backpropagation properties discussed before. It is worth noting that compared to plane DRNN, deep residual RNN adds neither extra parameters nor computational complexity (other than simple additions) and still aids in better classification performance.

Table 3.4: Performance comparison in terms of the mean±standard deviation of the metrics for the plane DRNN and the deep residual RNN models (10 runs).

Method	**Se**(%)	**Sp**(%)	**Acc**(%)	**AUC**
Plane DRNN	87.74 ± 9.89	86.87 ± 15.20	87.42 ± 11.04	0.87 ± 0.11
Deep residual RNN	**93.18** ± **7.10**	**94.05** ± **5.61**	**93.81** ± **3.84**	**0.93** ± **0.04**

Table 3.5: Performance comparison in terms of the mean±standard deviation of the metrics for the deep residual RNN without and with attention module (10 runs).

Method	**Se**(%)	**Sp**(%)	**Acc**(%)	**AUC**
Proposed method without attention	93.18 ± 6.10	94.05 ± 5.61	93.81 ± 3.84	0.93 ± 0.04
Proposed A-DRRNet model	**99.01** ± **0.58**	**97.89** ± **1.54**	**98.57** ± **0.99**	**0.98** ± **0.01**

3.2.3.2 Significance of the Attention Module

To evaluate the effectiveness of incorporating the attention module, we perform additional experiments by removing the attention module from the proposed A-DRRNet model. Table 3.5 compares the proposed method without and with the attention module. As can be seen that the proposed A-DRRNet model with attention outperforms the model without attention with a minimum improvement of nearly 6% in Se, 4% in Sp, 5% in Acc, and 5% in AUC. Moreover, the proposed A-DRRNet shows less standard deviation (SD) in measures across the ten runs, thereby providing better generalization results. This improved performance demonstrates that the attention module aid in further mining discriminative features by attending to diagnostically prominent ECG features during the classification process.

In summary, from plane DRNN to the proposed method, there is a significant improvement of nearly 11% across all the measures (Tables 3.4 and 3.5). Therefore, the residual connections and the attention module individually improve the proposed method's classification performance.

3.2.3.3 Analysis of Diagnostic Transparency of the Method using the Learned Attention Weights

This section demonstrates the diagnostic transparency of the proposed method by visualizing the learned attention weights. Figure 3.6 (a) shows the mean heartbeat signals corresponding to the CHF and HCs classes with their error bands on the test data. As can be seen, the CHF class heartbeats have more variability in the ECG features over HCs. Figure 3.6 (a) also shows the attention weights heat-maps. The heat-maps highlight the prominent ECG features that contribute the most to identifying a particular class. As discussed previously, the pathological ECG patterns, including prolonged QRS- and QT-duration, low amplitude QRS-complexes, ST-segment elevation/depression, and deep S-waves, are often associated with CHF patients. Interestingly, the attention weights of the proposed method (see Figure 3.6 (a)) for the CHF and HCs demonstrate that the ECG region, i.e., the RS- and ST-segments, which includes the aforementioned critical ECG patterns, is consistently focused during the diagnosis process. To corroborate the above discussion, Figure 3.7 presents the attention weights superimposed on the typical ECG beats of CHF and HCs. As observed, the method strongly focuses on diagnostically prominent ECG features, i.e., the RS-and ST-segments, which are critical for the correct detection. This visualization of attention heat-maps reveals the transparent diagnostic basis of the method, and it can help researchers and healthcare providers understand how the diagnosis decisions are reached. Specifically, for the continuous patient monitoring in remote-care or hospital-care applications using long-term (24 h) ECG signals, the attention heat-maps can help clinicians quickly examine the abnormal ECG patterns focused by the model than inspecting complete ECG recording.

3.2.4 Performance Comparison with the Existing Methods

To verify the effectiveness of the proposed method, we compare it with the following existing methods.

- *HRV features based ML method:* The proposed ECG-based diagnosis method is compared with the popularly used HRV features-based ML method. Most existing CHF detection methods employed the time- and frequency-domain HRV-based features computed from the RR-intervals to characterize the ANS disorder in the CHF patients. The popular time-domain indices include the SDNN, the RMSSD, and the pNN50. The frequency-domain indices include the LF, the HF, and the LF/HF. These HRV indices are computed from the different numbers of RR-intervals, i.e., the segments of 500

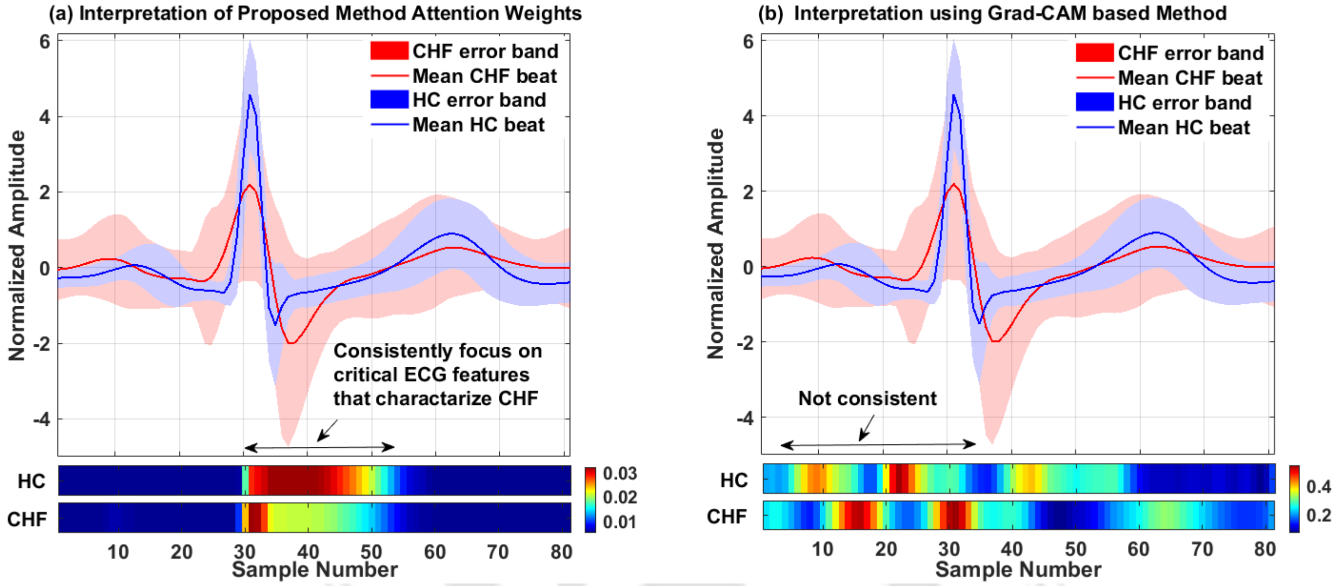


Figure 3.6: Comparison of model transparency results for (a) the proposed A-DRRNet method and (b) the Grad-CAM-based method [49]. In both (a) and (b), the mean CHF and HC heartbeats and their error bands are computed from the same test data.

RR-intervals (nearly equal to 5 min with 100 beats per minute), the segments of 100 RR-intervals, and the segments of 10 RR-intervals. The extracted features are provided to the RF classifier (number of estimators = 8) for obtaining the class predictions.

- *State-of-the-art DL-based CHF diagnosis methods using ECG heartbeat input:* Recent state-of-the-art DL-based CHF diagnosis methods, including the deep CNN (DCNN) developed by Acharya *et al.* [88], the CNN-Grad developed by Porumb *et al.* [49], and the CNN-RNN developed by Li *et al.* [89] are used for comparison. The DCNN [88] and CNN-Grad approaches use a stack of CNN layers to learn discriminative features from the raw ECG beats for the CHF diagnosis. Specifically, the CNN-Grad [49] also employed the Grad-CAM technique to demonstrate the model interpretability with the diagnosis. The CNN-RNN [89] utilizes CNNs for feature extraction, followed by RNNs for feature summarization to generate the diagnosis. These approaches were originally proposed to diagnose severe CHF patients (NYHA III and IV) and are adapted for the current dataset consisting of all stages of CHF patients (NYHA I-IV). For a fair comparison, we maintained the same data processing, splitting, and training iterations as the proposed method for the existing methods.

Table 3.6 compares the classification results for the proposed and existing methods. As can be seen, the feature-based method, i.e., the HRV+RF, provides reasonably good results with the long-term segments of 500 RR-intervals as input. However, the method's performance significantly degrades with the smaller RR-

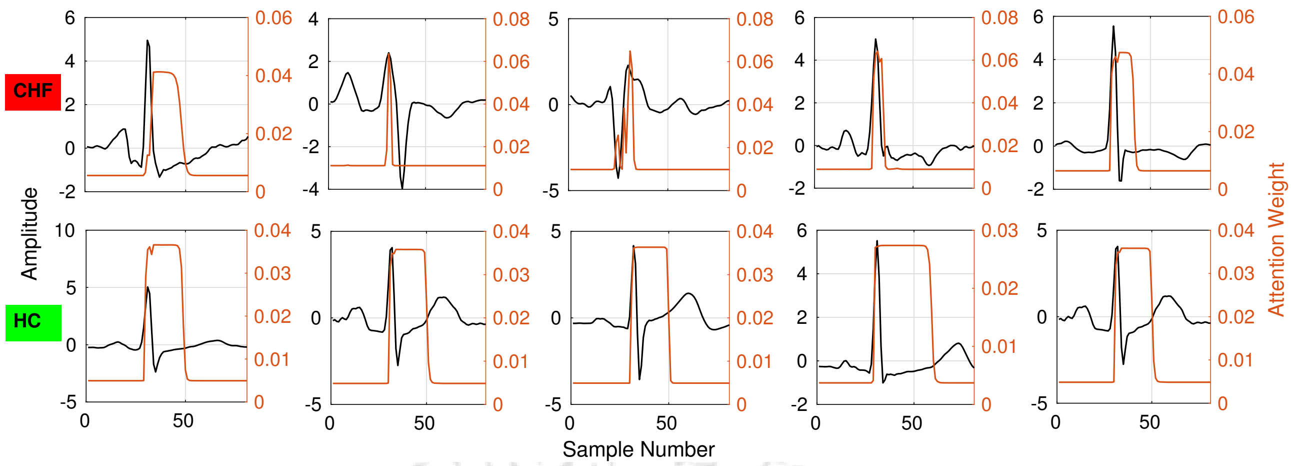


Figure 3.7: Proposed method's diagnostic transparency results for typical CHF and HC heartbeat profiles.

Table 3.6: Performance comparison of the proposed and the existing methods on the current database (10 runs).

Method	Input	Features	Classifier	Se(%)	Sp(%)	Acc(%)	AUC
HRV+RF	RR-intervals (500)	HRV features	Random forest	96.58 ± 3.11	99.44 ± 0.29	97.78 ± 1.24	0.98 ± 0.01
HRV+RF	RR-intervals (100)	HRV features	Random forest	91.31 ± 7.23	96.24 ± 2.24	93.12 ± 4.41	0.93 ± 0.02
HRV+RF	RR-intervals (10)	HRV features	Random forest	88.64 ± 7.36	95.18 ± 4.14	91.65 ± 7.18	0.91 ± 0.04
DCNN [88]	ECG beat	End-to-end CNN	Softmax	88.01 ± 9.76	96.99 ± 3.23	92.10 ± 5.41	0.92 ± 0.04
CNN-Grad [49]	ECG beat	End-to-end CNN	Softmax	95.96 ± 4.49	92.10 ± 7.02	93.94 ± 2.51	0.93 ± 0.02
CNN-RNN [89]	ECG beat	End-to-end CNN+RNN	Softmax	92.48 ± 6.31	97.67 ± 2.49	94.81 ± 3.17	0.95 ± 0.02
Proposed	ECG beat	End-to-end A-DRRNet	Softmax	**99.01 ± 0.58**	**97.89 ± 1.54**	**98.57 ± 0.99**	**0.98 ± 0.01**

intervals segments, limiting its applicability in mobile medical ECG devices. The end-to-end DL methods, i.e., the DCNN [88] and the CNN-Grad [49], fail to recognize the underlying temporal dynamics of diverse pathological ECG characteristics and tend to show poor classification results. Thus, the CNN-based models with local convolution filters are not optimal for representing complex ECG features concerning diverse CHF patients. The hybrid CNN-RNN [89] network shows a slight improvement in the Acc; however, the Se for CHF detection is only 92.48%. The existing methods show high SD in the measures, thereby demonstrating the poor generalization results as these methods fail to learn the discriminative features across various subjects. However, the proposed method outperforms existing methods by achieving an improved Se, Sp, Acc, and AUC of 99.01%, 97.89%, 98.57% and 0.98, respectively. The improved classification performance demonstrates the effectiveness of the proposed A-DRRNet for capturing temporal dynamics and extracting high-level attentive representations by focusing on clinically relevant ECG information. Moreover, the proposed method shows less SD in measures, rendering better generalization for the unseen patients. It is worth emphasizing that the HRV-based methods require long-duration RR-interval segments to achieve good performance. However, the proposed ECG waveform-based method generates improved results even with the heartbeat input by exploiting morphological ECG information.

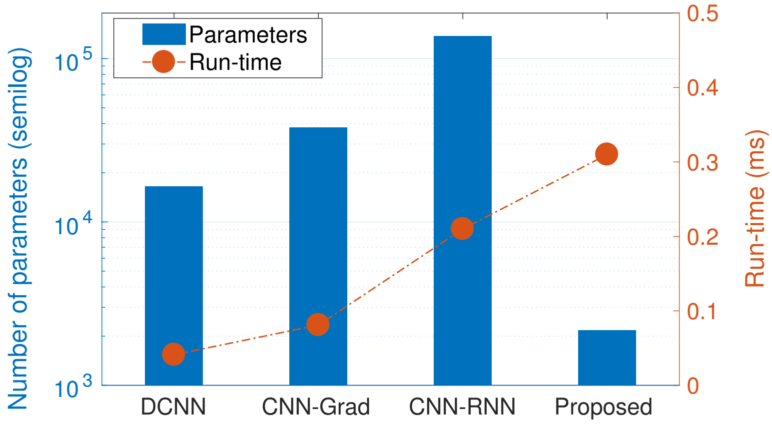


Figure 3.8: Comparison of number of trainable parameters and average test run-time of the proposed and existing DL methods.

Figure 3.8 compares the number of trainable parameters and the average test run-time of the proposed and existing DL-based methods. An i5 processor system with 16GB RAM and Quadro K600 graphics processor is used to perform the experiments. As observed, compared to the existing methods, the proposed method requires fewer training parameters, which alleviates the risk of over-fitting and helps improve classification performance. Although we designed the proposed model in a lightweight manner (fewer parameters), the sequential computation in the RNN layers slightly increases the run-time over the CNN-based methods. However, the test run-time of 0.3 ms to provide the classification decision is not significantly high to delay diagnosis. The proposed lightweight method with fewer training parameters, acceptable test-run time, and impressive classification results makes it suitable for mobile medical and remote healthcare systems using portable ECG devices.

Importantly Figure 3.6 compares the model transparency results for the proposed and the Grad-CAM-based method [49] for CHF detection. As can be seen from the average heat-maps of the Grad-CAM method (Figure 3.6 (b)), the highlighted ECG features that are class discriminative are inconsistent across the HCs and CHF classes, thus failing to incorporate the clinician's aspects of CHF diagnosis. Specifically, the highlighted features lack correlation with the ECG features that characterize the CHF, which results in a sub-optimal classification Acc of 93.94% (Table 3.6). However, the proposed method consistently focuses on the diagnostically prominent ECG region (see Figure 3.6 (a)), i.e., RS- and ST-segments, which are crucial to capture prolonged and low-amplitude QRS-complex, deep S-wave, J-point and ST-elevation/depression that characterize the CHF patients. The significant difference in both methods' results can be attributed to the fact that, unlike the Grad-CAM method, the proposed method enables the network to automatically focus on the diagnostically relevant ECG features in the training phase through the attention module. The promising model transparency results can advance the research towards interpretable models for CHF detection and improve the applicability of DL methods in the healthcare domain.

3.3 Summary

The preliminary screening of CHF patients using ECG signals helps identify suspected individuals who can benefit from further examination with an advanced echocardiogram (ECHO) for appropriate management. To this end, in this chapter, an attention-based deep residual RNN is presented for the automated diagnosis of CHF from the ECG signals. The method systematically processes the input ECG with the deep residual RNNs followed by an attention module to capture the temporal dynamics and extract high-level attentive representations effectively for the improved diagnosis. The experimental results verify that the proposed method outperforms the existing methods. A notable advantage of the method is the diagnostic transparency provided by the attention weights. The visualization of attention heat-maps demonstrates that the proposed method automatically identifies the diagnostically prominent ECG features and provides them more weight during the diagnosis prediction. This model transparency is a crucial contribution to the clinical context that can improve medical experts' trust in deploying DL frameworks in healthcare applications. In summary, the proposed method meets three crucial aspects of the DL-based automated diagnosis system, i.e., high diagnostic accuracy, lightweight, and diagnostic transparency; thus, it is suitable for mass screening of CHF patients in clinics and hospitals using mobile and portable ECG devices.

4

Atrial Fibrillation Burden Estimation from Long-Term Single-Lead ECG using Multi-Task Deep CNN

Contents

Atrial fibrillation (AF) is the most common heart rhythm disorder that can progressively worsen over time [30]. AF can be intermittent and asymptomatic; as a result, it often remains undiagnosed until it causes a life-threatening condition, such as stroke and systematic thromboembolism [31, 154]. Notably, because of the intermittent or paroxysmal nature of the AF condition, it is highly likely to be undetected with the conventional snapshot 30 s ECG evaluation performed during routine hospital visits [31]. It is thus a disorder that would benefit from long-term continuous monitoring with Holter ECG devices. Long-term ECG monitoring can help improve the AF diagnosis (presence or absence) and facilitate quantifying AF condition as a spectrum of atrial disease severity known as “AF burden”. The AF burden is defined as the percentage of time an individual is in AF rhythm during a long enough monitoring period [136]. A greater AF burden is a sign of severe atrial myopathy and is associated with an increased risk for stroke and reduced health-related QoL [136]. Recent studies [134–136, 155] demonstrate that AF burden measure can provide improved stroke risk-stratification and may help guide appropriate personalized stroke prevention strategies. With the recent developments in wearable cardiac health monitoring devices such as Zio Patch, CardioSTAT, and Nuvant, it has become simple yet effective to collect long-term continuous ambulatory ECG recordings [32]. The accurate estimation of AF burden from these recordings prompts an improved diagnosis, phenotyping, and management of AF in the context of remote health monitoring applications.

Existing ML-based [91–98, 100–112] and DL-based [90, 113–118] methods focus on detecting the presence or absence of AF using snapshot ECG and do not quantify the disease condition through AF burden measure using long-term ECG recordings. Recently, Chocron *et al.* [156] have developed a hybrid model combing CNNs, and recurrent neural networks (RNNs) named ArNet for AF burden estimation from the long-term recordings using RR-interval series data. However, this rhythm-based approach has the following limitations for precise AF burden estimation. (i) AF is not the only arrhythmia characterized by irregular RR-intervals. For example, frequent ectopic beats, i.e., premature atrial and ventricular contractions (PACs/PVCs) will also have irregular beat patterns typical of AF and are known to increase false AF detection rates, thereby increasing error in AF burden estimation [90]. (ii) The ECG signals acquired during ambulatory monitoring are prone to different noises, which degrade the R-peak detection performance and typically lead to inaccurate AF burden estimation.

To address these issues, in this chapter, we propose a novel multi-task deep convolutional neural network (MT-DCNN) approach that aims at accurate and robust AF diagnosis and its burden estimation from the ECG segments of long-term ambulatory recordings. In general, the supervised end-to-end learning approaches [90, 116–118] exploit the ECG data and their corresponding labels (AF or non-AF) to learn feature representations that improve classification performance. On the other hand, unsupervised (ECG data without class labels) frameworks such as convolutional denoising autoencoders (CDAE) composed of encoder and decoder structures have shown their effectiveness in reconstructing the denoised output ECG sequence from the noisy input ECG. The encoder part of CDAE learns a compact and robust representation of ECG sequence that can be used for many downstream classification tasks [157]. Therefore, because the long-term ECG recordings are prone to different noise levels and intra- and inter-patient variabilities, a multi-task model combining the ECG classification task and the ECG sequence reconstruction task may take advantage of each approach to lean robust feature representations for improving AF diagnosis and its burden estimation performance. Inspired by this idea, in this chapter, we present an MT-DCNN model consisting of AF detection as a primary task and reconstruction of ECG sequence as an auxiliary task using DCNNs. The auxiliary task regularizes the model to learn robust feature representations for efficient AF detection, thereby aiding accurate AF burden estimation. We demonstrate that the analysis of AF burden from long-term recordings enables improved diagnosis and characterization of AF condition.

The rest of the chapter is organized as follows. Section 4.1 and 4.2 present the details of the clinical ECG database and the proposed MT-DCNN model for AF diagnosis and its burden estimation, respectively. Section 4.3 presents the comprehensive and rigorous benchmark of the proposed MT-DCNN model against baseline AF detectors on four diverse PhysioNet datasets. The discussion of the proposed approach is presented in section 4.4, followed by the summary of the chapter in section 4.5.

4.1 Clinical ECG Database and Preprocessing

Four PhysioNet long-term ECG databases [148], namely the Long-Term Arial Fibrillation Database (LTAFDB), the MIT-BIH Atrial Fibrillation Database (AFDB), the MIT-BIH Normal Sinus Rhythm Database (NSRDB), and the MIT-BIH Long-Term Normal Sinus Rhythm Database (LTNSRDB), totaling n=134 patients and t=2,661 h of long-term ambulatory ECG data is used for experimentation. The details are as follows.

4.1.1 Long-Term Atrial Fibrillation Database

This database [148] consists of ambulatory ECG recordings of n=84 subjects suffering from paroxysmal AF. Each ECG record sampled at 128 Hz, with 12-bit resolution over a 20 mV range for about 22.7$\pm$2.4 h long. The overall database totals t=1,900 h, including 874 h spent in AF and 1,026 h in non-AF rhythms. The database is composed of various rhythm annotations, including normal rhythm (N), atrial fibrillation (AF), ventricular bigeminy (B), supraventricular tachyarrhythmia (SVTA), ventricular trigeminy (T), idioventricular rhythm (IVR), sinus bradycardia (SBR) and atrial bigeminy (AB). In the present analysis, we used AF rhythm type as “AF” class and all other rhythm types as “non-AF” class.

4.1.2 MIT-BIH Atrial Fibrillation Database

This database [148] consists of n=25 long-term, 10 h ambulatory ECG recordings from the patients with paroxysmal AF, totaling t=229 h of data, with 92 h spent in AF and 137 h in non-AF rhythms. Each ECG record sampled at 250 Hz, with 12-bit resolution, over a range of $\pm$10 mV. The database is composed of various rhythm annotations including atrial fibrillation (AF), atrial flutter (AFL), junctional rhythm (J), and other rhythms. In the present analysis, we used AF rhythm type as “AF” class and all other rhythm types as “non-AF” class. Two individuals are excluded from the study as they do not have ECG data.

4.1.3 MIT-BIH Normal Sinus Rhythm Database

This database [148] consists of n=18 long-term, 21.2$\pm$1.2 h long, two-channel ECG recordings, totaling t=384 h of data sampled at 128 Hz.

4.1.4 MIT-BIH Long-Term ECG Database

This database [148] consists of n=7 long-term, 21.1$\pm$3.4 h long, two-channel ECG recordings, totaling t=148 h of data sampled at 128 Hz.

Subjects of both NSRDB and LTNSRDB datasets had no serious arrhythmias apart from the presence of few ectopic beats. Thus, all the recordings from these databases are labeled as “non-AF” rhythm type. From all the databases, first channel ECG recordings are extracted and used for the analysis. ECG signals from the AFDB database are downsampled to 128 Hz to maintain a uniform sampling rate across all datasets. Current guidelines define the presence of AF as ECG documentation of absolutely irregular RR-intervals and no discernible, distinct P-waves lasting for at least 30 s [31, 32]. A significant factor in selecting this 30 s duration is that the automated AF diagnosis methods demand a certain number of RR-intervals to assess variability with sufficient confidence to trigger the diagnosis. In line with clinical guidelines, the long-term ECG recordings are divided into 30 s excerpts and used for the analysis. Each excerpt is labeled (AF/non-AF) according to their rhythm annotations. Table 4.1 summarizes the specific details of four databases in terms of total patients and the number of extracted AF/non-AF segments. Table 4.2 compares the proportion of ectopic (PACs/PVCs related) beats present across different databases. As can be seen, the LTNSRDB and the LTAFDB datasets contain a significant number of ectopic beats that may be useful to assess false positive AF detection rates. Recordings of AFDB database did not have any beat annotation files to assess the ectopic beat frequency.

In the present research, the LTAF database is used as the development dataset since it contains a relatively large number of AF and non-AF rhythm excerpts (Table 4.1) and thus can be used to determine

Table 4.1: Details of four PhysioNet long-term ECG datasets used in this study.

Database	Total patients	#Low-risk AF patients	#High-risk AF patients	#Non-AF patients	#AF segments	#Non-AF segments
LTAFDB	84	17	67	N/A	126,550	107,768
AFDB	23	5	18	N/A	11,364	16,739
NSRDB	18	N/A	N/A	18	N/A	52,488
LTNSRDB	7	N/A	N/A	7	N/A	17,694
Total	132	24	83	25	137,914	194,689

#: Number of; N/A: Not applicable.

Table 4.2: Comparison of ectopic beat frequency across the datasets.

Database	Number of patients	Total number of beats	Number of PAC/PVC related beats
LTAFDB	84	8,995,658	284,785 (3.17%)
AFDB	23	1,122,453	N/A
NSRDB	18	1,742,340	127 (0.007%)
LTNSRDB	7	761,528	68,608 (9.01%)

N/A: Not available.

stable model parameters. One of the critical aspects to consider while evaluating an automated AF detection model is its ability to generalize on external or independent test sets (not used in training) derived from different population samples and recorded with different acquisition devices. Therefore, three PhysioNet databases, including AFDB, NSRDB and LTNSRDB, are used as the external test datasets to assess the model generalization performance.

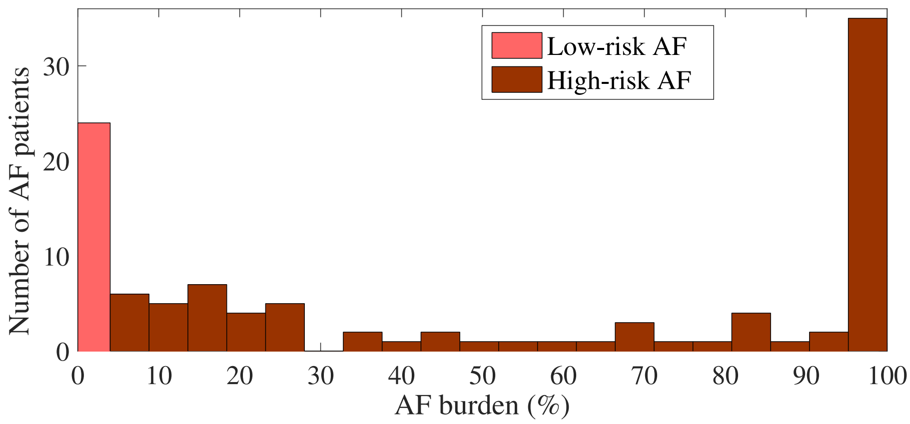


Figure 4.1: Histogram of AF burden for the low-risk and the high-risk AF individuals. This figure emphasizes the diversity of AF phenotypes in terms of AF burden measure for paroxysmal atrial fibrillation patients (n=107).

4.1.5 Dataset for AF Risk-Stratification

Recent studies [135, 136] have demonstrated that a maximum daily burden of longer than one hour, i.e., greater than 4%, carries important negative prognostic implications and may be a clinically relevant parameter for stroke risk-stratification in AF patients. Following these research recommendations, we defined an overall rhythm label for each patient as follows; i) non-AF: total time spent in AF below 30 s, ii) low-risk AF: total time spent in AF above 30 s and AF burden below 4% and iii) high-risk AF: AF burden above 4%. Table 4.1 shows the number of patients belonging to each of the defined groups across four databases. Figure 4.1, shows the distribution of AF burden for low-risk and high-risk AF patients. It emphasizes that the AF is not a binary entity but a spectrum of atrial disease severity characterized by AF burden. Thus, accurately estimating the AF burden is crucial to individualize a patient's risk for stroke.

4.1.6 Preprocessing

The raw ECG recordings are first divided into 30 s non-overlapping excerpts, and then each excerpt is processed to remove baseline wander noise using a second-order Butterworth low-pass filter with a cut-off frequency of 0.5 Hz. The ECG excerpts of long-term ambulatory recordings are often corrupted with high-frequency noises such as muscle contraction and electrode motion noises, which severely obscure the morphological ECG characteristics. In order to discard such clinically not useful excerpts, we employed an R-peak detection-based signal quality index (bSQI) (see Appendix A). Following the signal quality guidelines form [8, 156], excerpts with bSQI$<$0.8 are excluded from the current analysis. In addition, to equalize the signal amplitude across recordings and databases, each ECG excerpt is normalized to the interval [0,1] using min-max normalization (see Appendix A).

4.2 Multi-Task Deep CNN Approach for AF Diagnosis and AF Burden Estimation

This section presents the proposed multi-task deep convolutional neural network (MT-DCNN) model that combines the ECG classification and sequence reconstruction tasks into a single framework for the accurate and robust AF diagnosis and its burden estimation. Specifically, the MT-DCNN incorporates ECG reconstruction as an auxiliary task to improve the feature extraction pipeline for the primary AF/non-AF classification task. Instead of learning features for classification only, the MT-DCNN model also tries to represent morphological features of the ECG through sequence reconstruction. Thus, the auxiliary task acts as a regularization term for the classification and prevents it from overfitting, thereby learning robust and generalizable feature representations. The two tasks shared a common network structure in part, and features are learned using a combined loss. The block diagram of the method is shown in Figure 4.2. As can be seen, the MT-DCNN model consists of three main modules: the shared encoder, the classifier, and the decoder. The details of each of the modules are described as follows.

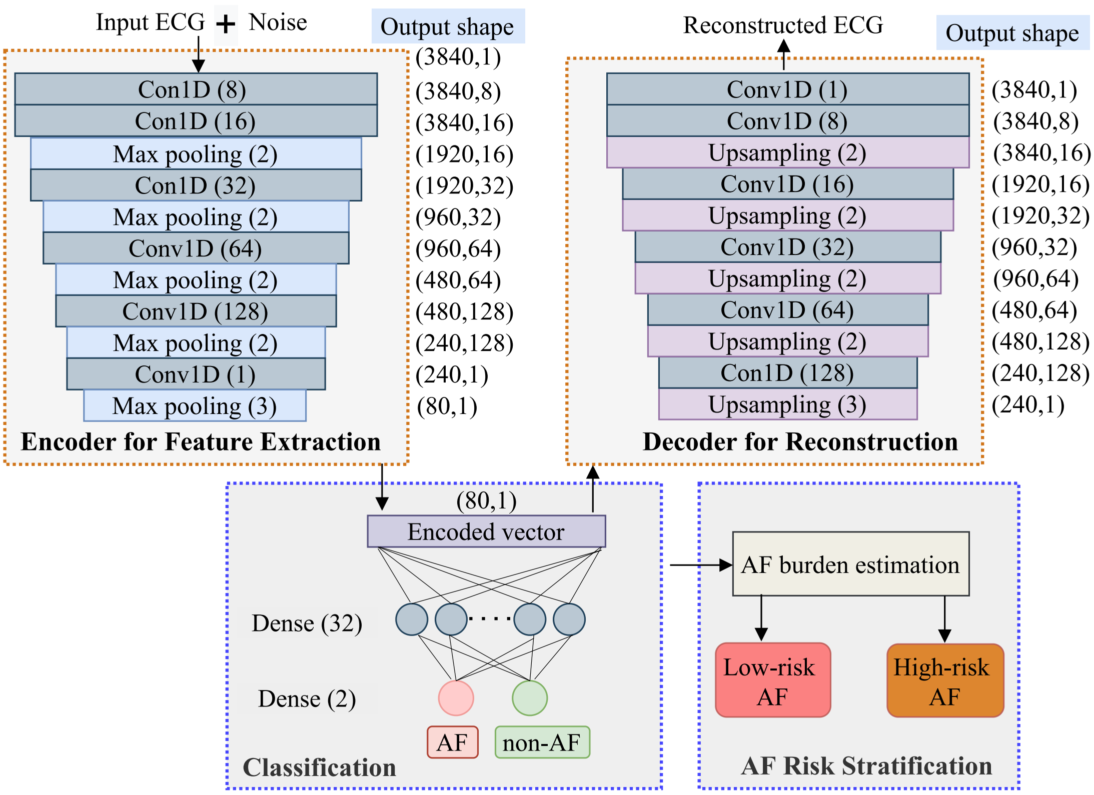


Figure 4.2: The MT-DCNN architecture illustrates the data flow and the output data shape at each layer. The model process the ECG segments of long-term ambulatory recordings and classify them as either AF or non-AF rhythms. An auxiliary task of ECG sequence reconstruction using a convolutional denoising autoencoder structure is incorporated into the model to improve the feature extraction pipeline. The predicted AF/non-AF decisions are used to estimate the AF burden measure. This measure is further analyzed to stratify AF individuals at high- and low-risk for stroke.

4.2.1 Shared Encoder Module

In order to learn robust feature representations, the original input ECG is corrupted with Gaussian noise due to its resemblance with high-frequency noises of ECG and fed to the encoder as an input [157]. The encoder module consists of six 1D-convolutional layers and five pooling layers to learn compact feature representations from the high-dimensional input ECG signals (Figure 4.2). All the convolutional layers will have filters with a fixed size of 1×7 and all the local max-pooling layers will have a filter size of 1×2, except the last pooling layer with filter size 1×3. The original input ECG is denoted as $\mathbf{x} \in \mathbb{R}^T$ and the noise corrupted input ECG as $\tilde{\mathbf{x}} \in \mathbb{R}^T$ with T represents the length of input sequence. Encoder module process $\tilde{\mathbf{x}}$ and learn a compact encoded representation $\mathbf{z} \in \mathbb{R}^{T'}$ with T' denotes dimension of the encoded vector.

4.2.2 Classification Module

The encoded vector $\mathbf{z}$ is connected to two fully connected (FC) layers; where the last FC layer is with a Softmax activation to output the probability of AF or non-AF class label for each ECG excerpt.

4.2.3 Decoder Module

The decoder network is used to reconstruct the original ECG sequence based on the compact encoded vector $\mathbf{z}$ of the encoder module. As shown in Figure 4.2, the decoder module is systematically designed using five upsampling layers and six 1D-convolutional layers to output the reconstructed ECG sequence $\hat{\mathbf{x}} \in \mathbb{R}^T$ with the same size T as input ECG $\mathbf{x}$. All the upsampling layers will have a size of 1×2, except the first upsampling layer with a size of 1×3 and all the convolutional layers will have filters with a size of 1×7.

4.2.4 Multi-Task Learning

The proposed MT-DCNN model with a shared encoder combines the classification task and the ECG sequence reconstruction task into a single framework (Figure 4.2). In order to learn robust encoded feature representations, a combined loss function with classification and reconstruction objectives is used to update the MT-DCNN model parameters. The combined loss function (L) is formulated as follows:

$$L = L_c^{ECG} + \lambda L_r^{ECG}, \lambda \geq 0 \tag{4.1}$$

where L_c^{ECG} and L_r^{ECG} represents the classification loss and the reconstruction loss, respectively. λ is a constant factor weighing losses and is set to one via experimentation (see Table 4.6). For a model training for the classification task (AF or non-AF), the objective is to minimize the binary cross-entropy (BCE) loss computed between the true class label y and the predicted label $\hat{y}$ for each sample excerpt given as:

$$L_c^{ECG} = -\big(ylog(\hat{y}) + (1-y)log(1-\hat{y})\big) \tag{4.2}$$

Similarly, for the ECG sequence reconstruction task, the objective is to minimize the mean squared error (MSE) between the original input ECG ($\mathbf{x} \in \mathbb{R}^T$) and the reconstructed output ECG ($\hat{\mathbf{x}} \in \mathbb{R}^T$) for each excerpt computed as:

$$L_r^{ECG} = \frac{\|\mathbf{x} - \hat{\mathbf{x}}\|^2}{T} \tag{4.3}$$

where T represents the sequence length. The average combined loss (L) over a set of training samples is used to update all the trainable parameters of the MT-DCNN model simultaneously. All the convolutional layers in the encoder and decoder are followed by a batch normalization (BN) layer with a ReLU activation. However, the last convolutional layer in the decoder will have a Sigmoid activation function for reconstructing ECG sequence in the range [0,1].

4.2.5 AF Burden Estimation and Stroke Risk-Stratification

The MT-DCNN model generates AF/non-AF diagnosis for each ECG excerpt of long-term ambulatory recordings. These diagnosis decisions are used to estimate AF burden (percentage of time in AF rhythm) for the long-term recordings as [156]:

$$AF\ burden = \left[\frac{\sum_{i=1}^{N} t_i \times I_i}{\sum_{i=1}^{N} t_i}\right] \times 100 \tag{4.4}$$

where t_i represents the length of i^{th} ECG excerpt in seconds, N is the number of excerpts and I_i is an identity operator which is equal to one when the excerpt have AF rhythm and zero otherwise. Following the recent research guidelines [135, 136, 156], the estimated AF burden is further analyzed to identify patients who are at substantially high-risk for stroke and relatively low-risk for stroke.

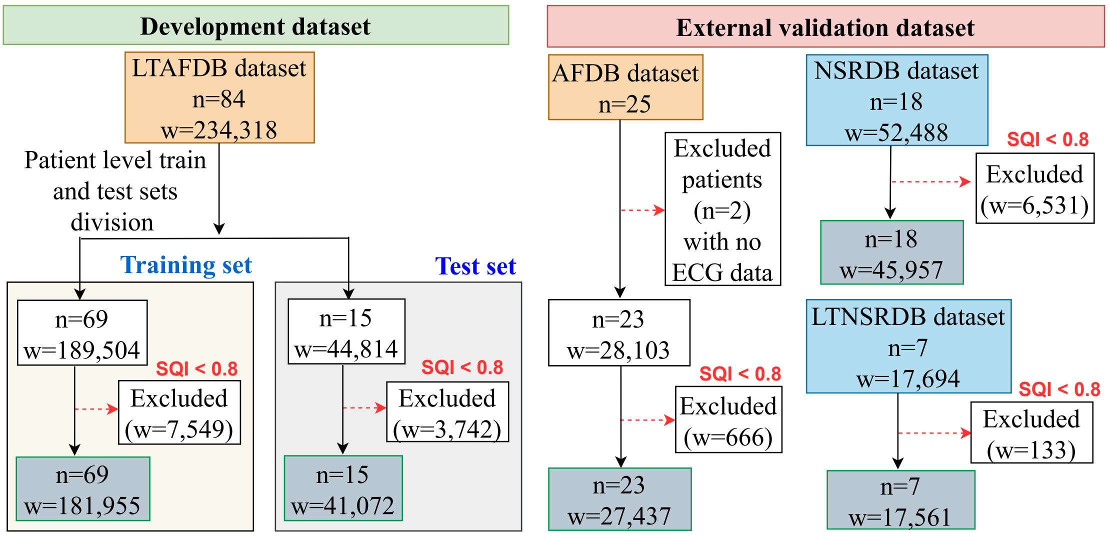


Figure 4.3: The ECG data selection process applied to the development dataset (LTAFDB) and the independent or external validation datasets (AFDB, NSRDB and LTNSRDB). Here, SQI denotes signal quality index; n denotes patients; and w denotes windows or excerpts.

4.3 Experimental Results

This section presents the details of the data selection, performance measures, and training settings, followed by the comprehensive and rigorous benchmark of the proposed MT-DCNN model against baseline AF detectors on four PhysioNet datasets.

4.3.1 Data Selection

From the four PhysioNet databases, we have extracted n=132 long-term recordings composed of t=2,661 h of ambulatory ECG data. Table 4.1 shows the extracted excerpts for the AF and non-AF rhythms. Table 4.2

shows the ectopic beat frequency across the four datasets, which helps to assess the effect of frequent ectopic beats on the AF detector's performance. Figure 4.3 illustrates the data selection process for the model development (LTAFDB) and independent validation (AFDB, NSRDB and LTNSRDB) datasets. As can be seen, excerpts with bSQI less than 0.8 are discarded due to severely poor quality. Specifically, for the LTAFDB, the quality check resulted in w=223,027 excerpts suitable for the analysis, including 122,034 (54.7%) AF and 100,993 (45.3%) non-AF excerpts. These excerpts are divided into train and test sets in a patient-independent manner composed of w=181,955 (AF: 93,746, non-AF: 88,209) and w=41,072 (AF: 24,258, non-AF: 16,814) excerpts, respectively (Figure 4.3). Similarly, the quality check resulted in a total of w=27,437 (AF: 11,105, non-AF: 16,332) excerpts for AFDB, w=45,957 (non-AF) for NSRDB and w=17,561 (non-AF) for LTNSRDB databases.

4.3.2 Performance Measures

In order to assess the excerpt-level classification performance (AF/non-AF), the following overall measures are computed: sensitivity (Se), specificity (Sp), positive predictive value (PPV), accuracy (Acc), $F1$-score and normalized Matthews correlation coefficient (nMcc) [90]. In addition, to evaluate the accuracy of AF burden estimation for the long-term recordings, we employ an absolute AF burden estimation error measure (E_{AF}) as follows [156]:

$$E_{AF} = \frac{\sum_{i=1}^{N} t_i \times |y_i - \hat{y}_i|}{\sum_{i=1}^{N} t_i} \times 100 \tag{4.5}$$

where y_i represents the true rhythm label of the i^{th} excerpt (one for AF; zero for non-AF) and $\hat{y}_i$ represents the predicted rhythm label for the same i^{th} excerpt (one for AF; zero for non-AF). For each patient group (low-risk AF, high-risk AF and non-AF), five statistical measures including minimum, Q1, median, Q3 and maximum of E_{AF} are also reported.

4.3.3 Network Parameters and Training Settings

The MT-DCNN network is developed in Keras using the TensorFlow backend. An NVIDIA Tesla V100 GPU is used to perform the experiments. The convolution filter size of 1×7 and the FC layer size of 32 are optimized using a grid-search technique (see Figure 4.2). The FC layers are followed by a dropout layer with a drop probability of 0.3 to mitigate the risk of overfitting. The model is trained using an Adam optimizer with a learning rate of 0.001 on mini batches of size 32 for 100 epochs. An early stopping criterion is employed, which stops the training when the F1-score does not improve over 20 iterations. The best-performing model is used for the performance evaluation. The learning rate, drop rate, and batch size are set experimentally for stable and faster training of the MT-DCNN model.

Table 4.3: Rhythm-based and morphology-based features extracted for training expert-crafted AF detectors.

Category	Number	Feature	Definition
	1	CosEn	Coefficient of sample entropy [99].
	2-5	AFEv, PACEv, OriginCount, IrrEv	Measures derived from the Lorenz plot to access RR-interval irregularities [94].
Rhythm	6	minRR	The minimal RR-interval duration.
(ventricular	7	AVNN	The mean RR-interval over the segment.
activity)	8	SDNN	The standard deviation of the RR-intervals over the segment [156].
	9	RMSSD	The root mean square of the successive differences [96].
	10-11	pNN20, pNN50	The percentage of RR-intervals shorter than 20 ms and 50 ms, respectively [156].
	12	SCN	Spectral concentration of atrial activity signal [108].
Morphology	13	DF	Dominant frequency of atrial signal [107].
(atrial	14	fWP	f-wave power in ECG [107].
activity)	15	fWPMAW	f-wave power in main atrial wave [107].
	16	SampEnfW	Sample entropy of atrial activity signal [106].

4.3.4 Baseline Methods used for Performance Comparison

Most existing methods focus on detecting the presence or absence of AF and do not analyze the AF burden using long-term ambulatory ECG recordings. Therefore, we benchmark the AF diagnosis and AF burden estimation performance of the following representative baseline AF detectors on four PhysioNet datasets and compare them with the proposed MT-DCNN model.

- *Expert-crafted AF detectors:*
  - *Rhythm-based AF detector:* From the literature, we identified eleven rhythm-based features (see Table 4.3) widely investigated for automated AF detection using the RR-intervals series. These features include Coefficient of Sample Entropy (CosEn) [99]; a set of features from the Lorenz Plot [94], i.e., AFEvidence (AFEv), Premature Atrial Complex Evidence (PACEv), Origin Count (OriginCount) and Irregularity Evidence (IrrEv); several time-domain statistical features such as minimal RR-interval duration (minRR), mean RR-interval over an excerpt (AVNN), the standard deviation of the RR-intervals over an excerpt (SDNN), root mean square of the successive differences (RMSSD) [96] and percentage of RR-intervals shorter than 20 ms and 50 ms (pNN20 and pNN50) [156].
  - *Rhythm- and morphology-based AF detector:* In addition to the rhythm-based features, few researchers have explored the atrial activity signal (residual signal after QRST cancellation) features. Table 4.3 shows four popular features extracted from the spectral analysis of the

atrial f-waves signal. Following features are extracted from the frequency power spectrum: dominant frequency (DF), defined as the frequency with the highest power between 3-12 Hz of the spectrum [107]; area under the power spectrum of atrial f-waves (fWP) [107]; spectral concentration (SC) defined as the ratio of spectral power in the 3-12 Hz band to the total power of atrial f-waves signal [108]; f-wave power in the main atrial wave (fWPMAW) [107]. Finally, Sample Entropy of the atrial f-waves signal (SampEnfW) is also employed for characterizing f-wave irregularity [106].

– All the features are standardized to have zero mean unit variance. In order to design a rhythm-based AF detector, an RF classifier is trained on rhythm-based features, and to design a rhythm- and morphology-based AF detector, an RF classifier is trained on both rhythm- and morphology-based features. These detectors return a rhythm label (AF/non-AF) for each ECG excerpt and are benchmarked on four PysioNet databases. We choose the RF classifier because it is an ensemble ML classifier widely explored in the literature studies [101, 102]. The hyperparameters of the RF classifier, such as the number of estimators (n_e=15) and the maximal depth of the trees (m_d=3), are optimized using five-fold cross-validation (CV) on the training dataset.

- *DL-based AF detector with RR-interval series as input:* This method uses DCNNs for automatically extracting features from the RR-interval series for AF diagnosis. The design of this DCNN rhythm-based model is inspired from [156], which used windows of 60 RR-intervals for training and composed of two convolution layers with filters 16, and 32 respectively, followed by a max-pooling layer of size 1×2, followed by one more convolution layer with 64 filters. All the convolutional filters had a fixed size of 1×10. Three FC layers follow the CNN layers with sizes 128, 64 and 32. Finally, an FC layer with Softmax activation is used to generate AF/non-AF classification.

- *DL-based AF detectors with raw ECG sequence as input:* Two single-task versions of the MT-DCNN model are also employed for the performance comparison. (i) A single-task classifier configuration of the MT-DCNN, i.e., encoder+classifier, is designed for the DCNN ECG-based AF/non-AF classification (DCNN-ECG-C). (ii) An unsupervised ECG sequence reconstruction using the CDAE model is trained independently. To build an AF vs. non-AF classification model from the pre-trained CDAE network, a fully connected output layer with Softmax activation is added on top of the encoder part of the CDAE (DCNN-ECG-R). For a fair comparison, the baseline DL models follow the same preprocessing and training procedure as the proposed MT-DCNN model.

Table 4.4: Performance comparison for AF vs. non-AF ECG excerpt classification on the LTAFDB test set and the independent validation datasets (AFDB, NSRDB, and LTNSRDB).

Method	LTAFDB (Test set)						AFDB						NSRDB†		LTNSRDB†	
	Se(%)	**Sp**(%)	**PPV**(%)	**Acc**(%)	**F1**(%)	**nMcc**(%)	**Se**(%)	**Sp**(%)	**PPV**(%)	**Acc**(%)	**F1**(%)	**nMcc**(%)	**Sp**(%)	**Acc**(%)	**Sp**(%)	**Acc**(%)
Rhythm	94.8	89.9	93.1	92.8	93.9	92.5	96.7	94.2	91.9	95.2	94.3	95.1	96.8	96.8	78.8	78.8
Rhythm and morphology	95.1	90.2	93.3	93.1	94.2	92.8	97.3	96.3	94.7	96.7	96.0	96.6	**97.8**	**97.8**	77.9	77.9
DCNN-rhythm	96.1	89.6	93.1	93.4	94.5	93.2	98.0	95.5	93.7	96.5	95.8	96.4	94.0	94.0	74.0	74.0
DCNN-ECG-R (single-task)	95.3	90.1	93.3	93.2	94.3	92.9	91.7	95.5	94.2	93.4	93.7	93.2	92.9	92.9	83.5	83.5
DCNN-ECG-C (single-task)	93.7	97.5	98.2	95.3	95.9	95.2	92.3	95.3	93.0	94.1	92.6	93.8	93.4	93.4	88.2	88.2
Proposed MT-DCNN	**96.5**	**97.9**	**98.5**	**97.1**	**97.5**	**97.1**	**98.8**	**98.0**	**97.1**	**98.3**	**97.9**	**98.2**	96.3	96.3	**92.9**	**92.9**

†: NSRDB and LTNSRDB datasets contain only non-AF rhythm excerpts; thus, we report only Se or Acc for these datasets.

Table 4.5: Performance comparison of absolute AF burden estimation error ($E_{AF}(\%)$) statistics for the LTAFDB test set and the independent validation datasets (AFDB, NSRDB, and LTNSRDB). Min: minimum, Q1: first quartile, Med: median or second quartile, Q3: third quartile, Max: maximum, and Mean: average.

Method	LTAFDB (Test set)						AFDB						NSRDB						LTNSRDB					
	Min	**Q1**	**Med**	**Q3**	**Max**	**Mean**	**Min**	**Q1**	**Med**	**Q3**	**Max**	**Mean**	**Min**	**Q1**	**Med**	**Q3**	**Max**	**Mean**	**Min**	**Q1**	**Med**	**Q3**	**Max**	**Mean**
Rhythm	0.0	0.4	1.4	6.9	40.8	7.1	0.0	0.1	3.0	7.6	24.1	4.6	0.0	**0.07**	0.7	3.3	27.9	3.2	1.2	2.8	4.6	17.1	99.7	21.2
Rhythm and morphology	0.0	0.5	1.9	4.6	41.3	6.8	0.0	0.4	1.9	4.4	21.4	3.3	0.0	**0.07**	**0.6**	**2.0**	**21.1**	**2.2**	1.5	**2.6**	5.1	24.6	99.6	22.1
DCNN-rhythm	0.0	0.2	1.5	4.9	42.4	6.5	0.0	0.3	1.3	5.5	14.1	3.5	0.0	0.2	1.7	6.5	47.2	5.9	2.2	3.1	10.3	35.7	99.7	25.9
DCNN-ECG-R (single-task)	0.0	0.5	1.8	4.6	40.3	6.9	0.0	0.3	4.8	12.9	27.7	6.6	0.0	1.8	3.6	6.9	50.3	7.1	1.1	5.9	9.8	17.9	91.7	16.5
DCNN-ECG-C (single-task)	0.0	0.03	0.7	3.1	25.6	4.6	0.0	1.1	2.3	10.6	15.2	5.9	0.0	1.2	2.5	7.5	49.5	6.6	0.6	11.6	13.0	16.3	18.7	11.7
Proposed MT-DCNN	0.0	**0.01**	**0.6**	**2.4**	**23.2**	**2.8**	0.0	**0.02**	**0.6**	**1.8**	**8.0**	**1.7**	0.0	1.1	1.6	3.6	30.1	3.7	**0.2**	3.4	**3.8**	**12.6**	**16.8**	**6.9**

4.3.5 Evaluation of the Proposed and Baseline AF Detectors Performance

In this section, the AF detectors are compared in terms of AF diagnosis and burden estimation performance.

4.3.5.1 AF Diagnosis Performance

Table 4.4 compares the excerpt classification results for the proposed MT-DCNN and baseline models. As seen for the LTAFDB test set, the MT-DCNN model outperformed all the baseline models with an overall nMcc of 97.1$\%$. In summary, deep learning models perform better than feature engineering methods. The proposed MT-DCNN model shows consistent improvement in all measures over other deep learning models. Table 4.4 also shows the model generalization results for three independent datasets (AFDB, NSRDB and LTNSRDB). As seen, the proposed model achieves an improved nMcc of 98.2$\%$ for the AFDB dataset. The NSRDB and LTNSRDB datasets contain only non-AF rhythm excerpts; thus, we report only Sp or Acc for these datasets. The proposed model achieves comparable Sp of 96.3$\%$ for the NSRDB and considerably improved Sp of 92.9$\%$ for the LTNSRDB datasets. Importantly, the AF detectors that depend only on RR-intervals information (the rhythm-based and the DCNN rhythm-based) show lower Sp values; due to lack of P-wave information for discriminating other rhythms with irregular RR-intervals from the AF rhythm. Even though rhythm- and morphology-based detector uses both RR-intervals and P-wave morphology information, the improvement in Sp measure is minimal. This is because the low-amplitude atrial activity signal features are commonly obscured by noise, leading to no significant performance gains. Similarly, the single-task DCNN-ECG-R model shows limited performance due to not exploiting the class labels during training; thus, the learned features are not optimized well for the AF/non-AF classification task. However, the proposed MT-DCNN model with raw ECG as input learned robust discriminative representations, thereby showing consistent improvement in all measures over the baseline models. Moreover, the five-fold cross-validation on the development dataset shows an average accuracy($\%$) of 96.8$\pm$1.6. The low standard deviation of 1.6$\%$ across five-folds for the accuracy demonstrates the stability and generalizability of the proposed MT-DCNN model.

4.3.5.2 AF Burden Estimation Performance

Table 4.5 compares the absolute AF burden estimation error statistics for the proposed MT-DCNN and baseline models. As seen, the E_{AF} median and interquartile range (Q3-Q1), on the LTAFDB test set, is 0.6 (2.4-0.01) for MT-DCNN, 1.4 (6.9-0.4) for rhythm-based model, 1.9 (4.6-0.5) for rhythm- and morphology-based model, 1.5 (4.9-0.2) for DCNN rhythm-based model, 1.8 (4.6-0.5) for DCNN-ECG-R single-task model and 0.7 (3.1-0.03) for DCNN-ECG-C single-task model. Moreover, the mean E_{AF} for the MT-DCNN model is 2.8, and it is significantly lesser (p<0.05) compared to the feature engineering models and other DL models. Generalization results on the independent datasets reveal that the MT-DCNN shows lesser (for AFDB and LTNSRDB) or comparable (for NSRDB) E_{AF} over the baseline methods. Specifically, the proposed model shows a mean E_{AF} of 1.7 on the AFDB, 3.7 on the NSRDB and 6.9 on the LTNSRDB. In order to evaluate the E_{AF} for a different group of individuals (non-AF, low-risk AF and high-risk AF); we pooled patients from all test datasets (LTAFDB test set: 15, AFDB: 23, NSRDB: 18 and LTNSRDB: 7) and created an overall test dataset of 63 individuals (25 non-AF, 7 low-risk AF, and 31 high-risk AF). The LTAFDB test set consists of 2 low-risk and 13 high-risk AF patients and the grouping details of other test sets are given in Table 4.1. Figure 4.4 shows the box and whisker plot illustrating the distribution of E_{AF} for the proposed method across different group of individuals. As observed, the median E_{AF} of the proposed MT-DCNN is considerably less across different groups. The inference time for a 24-hour long-term recording is 3.6 s for the rhythm-based, 3.8 s for the rhythm- and morphology-based, 1.4 s for the DCNN rhythm-based and 2.9 s for the proposed MT-DCNN model.

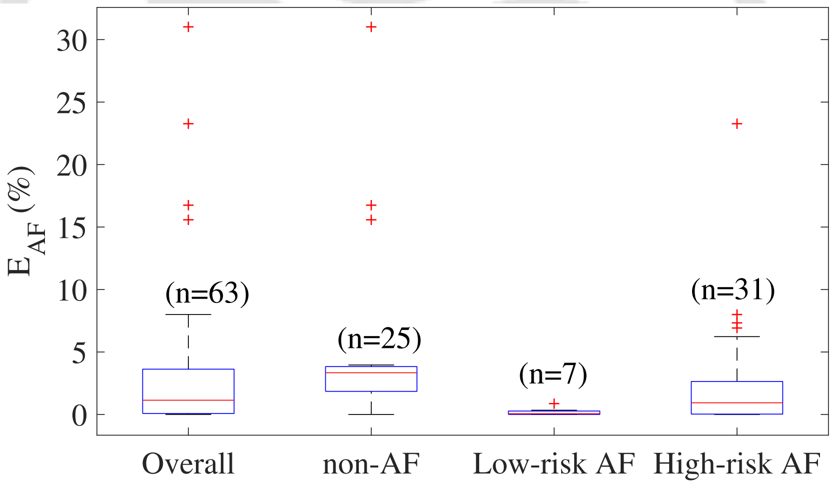


Figure 4.4: Boxplot of AF burden estimation error (E_{AF}) for the MT-DCNN method (at $\lambda = 1$) across different groups (all, non-AF, low-risk and high-risk) of individuals.

4.4 Discussion

Accurate AF burden estimation from the long-term ambulatory recordings can better characterize the AF condition and support improved clinical decision-making. In this study, we presented a novel multi-task deep learning model for precise AF burden estimation. The proposed model is developed and evaluated on a large LTAF database and the generalization of the trained model is verified on three independent datasets. The AF detection performance is evaluated in terms of Se, Sp, PPV, Acc, F1-score and nMcc. Specifically, following the evaluation recommendations from [90], nMcc is used to report the overall performance. In addition, to evaluate the accuracy of AF burden estimation, we defined an absolute AF burden estimation error measure (E_{AF}). The first observation from the experimental results (Table 4.5) is that it is possible to accurately estimate AF burden using single-channel long-term ECG recordings acquired via ambulatory monitoring. The proposed MT-DCNN model achieved the best performance in terms of E_{AF} on the LTAFDB test set. Moreover, the consistent performance on the three independent datasets demonstrates the model's generalization ability (Table 4.4 and Table 4.5).

For performance comparison, two expert-crafted (rhythm-based, rhythm- and morphology-based) and three DL-based AF detectors (DCNN rhythm-based, DCNN-ECG-C and DCNN-ECG-R) have been studied. Expert-crafted detectors are designed using several features extracted from the RR-intervals, the P-waves and the f-waves information (Table 4.3). The DL models automatically learn discriminative features from the RR-intervals (DCNN rhythm-based) or ECG sequence (DCNN ECG-based) data. The experimental results demonstrate that the proposed MT-DCNN outperformed the expert-crafted and DL-based detectors in terms of both AF detection (Table 4.4) and AF burden estimation performance (Table 4.5).

4.4.1 Effect of λ on the AF Burden Estimation Performance

The parameter λ is a constant factor weighing the auxiliary loss in the proposed combined loss function (Eq. 4.1). Table 4.6 shows the AF burden estimation error variation concerning λ on the development dataset. As can be seen, when the λ is 0, the MT-DCNN model is equivalent to the single-task classification model, which generates an average E_{AF} error of 4.6%. However, with increasing λ from 0, there is a significant reduction in the error. Specifically, at λ=1, the proposed MT-DCNN achieves the best performance in terms of lesser E_{AF} error.

Table 4.6: Variation of AF burden estimation error ($E_{AF}(\%)$) of proposed MT-DCNN model for different values of λ on the LTAF test set.

	$\lambda = 0$	$\lambda = 0.2$	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$	$\lambda = 1$	$\lambda = 1.2$	$\lambda = 1.4$
$E_{AF}(\%)$	4.6	3.9	3.8	3.6	3.1	**2.8**	2.9	3.2

4.4.2 Effect of Frequent Ectopic Beats on the AF Burden Estimation Performance

Ectopic beats (PACs/PVCs), known as AF-masquerading arrhythmias, are frequently detected on long-term ECG and are usually benign. The experimental datasets LTAFDB and LTNSRDB have a relatively high proportion of ectopic beats with 3.17% and 9.01%, respectively (Table 4.2). Thus, AF detection and its burden estimation on these datasets can be interpreted as performance as a function of ectopic beat frequency. As expected, methods that rely only on RR-intervals information (rhythm-based and DCNN rhythm-based) show poor Sp (Table 4.4) and higher AF burden estimation error (Table 4.5) as the ectopic beat rhythms are confused with AF, increasing false-positive rates. The performance degradation is more severe for the LTNSRDB dataset as it has a high frequency of ectopic beats. Since features from the low-amplitude f-waves were often affected by ambulatory conditions, its incorporation to the RR-intervals features did not yield considerable performance gains for the rhythm- and morphology-based model. The proposed single-task and MT-DCNN models with raw ECG sequence (atrial and ventricular information) input perform better over other baseline models on the LTAFDB test set. However, the single-task models show poor generalization results on the independent datasets (AFDB and NSRDB) compared to expert-crafted models (Table 4.4 and 4.5). This behavior is because the learned features are biased to the development dataset. However, by incorporating ECG sequence reconstruction as an auxiliary task, our proposed multi-task model, i.e., MT-DCNN, demonstrated improved AF detection and precise AF burden estimation results over baseline models. The reconstruction task served as implicit regularization for the classification task, which reduced overfitting in the classification problem. Hence, the generalization performance is improved for the AF detection (Table 4.4) and its burden estimation (Table 4.5).

4.4.3 Effect of Different Noise-Levels on the AF Burden Estimation Performance

In practice, the ECG signals acquired during ambulatory monitoring are often prone to high-frequency noises, which increases the false positive AF detection rates and thus affects the AF burden estimation accuracy. In order to investigate the robustness of proposed MT-DCNN and baseline models, their AF burden estimation error (E_{AF}) as a function of noise levels is presented in Figure 4.5. As expected, error of the rhythm-based, the rhythm- and morphology-based and the DCNN rhythm-based models increases rapidly at higher noise levels; because of error in QRS detector performance. In addition, for the rhythm- and morphology-based model, noise enters through P- and f-waves features, which further increased AF burden estimation error at higher noise levels. The proposed single-task and the MT-DCNN models consider the raw ECG sequence as input and do not depend on the QRS detection performance (Figure 4.5). Still, the structure of the single-task model is not robust at high noise levels and generated higher AF burden estimation error. However, with the incorporation of CDAE based ECG sequence reconstruction task, the proposed multi-task model generates significantly lower ($p<0.01$) AF burden estimation error over the baseline methods even at higher noise levels (see Figure 4.5). Thus, combining reconstruction and classification tasks into a single framework helps to learn compact and more robust feature representations, thereby achieving robust AF burden estimation performance.

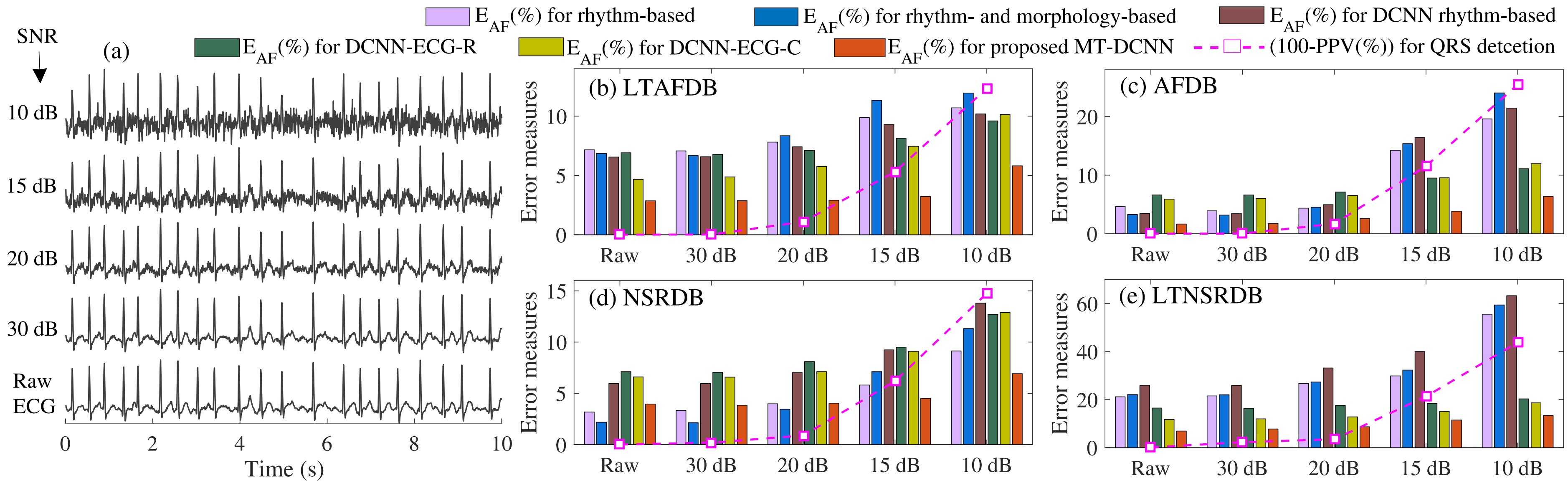


Figure 4.5: (a) A typical ECG excerpt with different noise levels in terms of SNR and (b)-(e) demonstrate mean absolute AF burden estimation error (E_{AF}) and the QRS detection error as a function of different noise levels on the LTAFDB test set, the AFDB, the NSRDB and the LTNSRDB datasets, respectively.

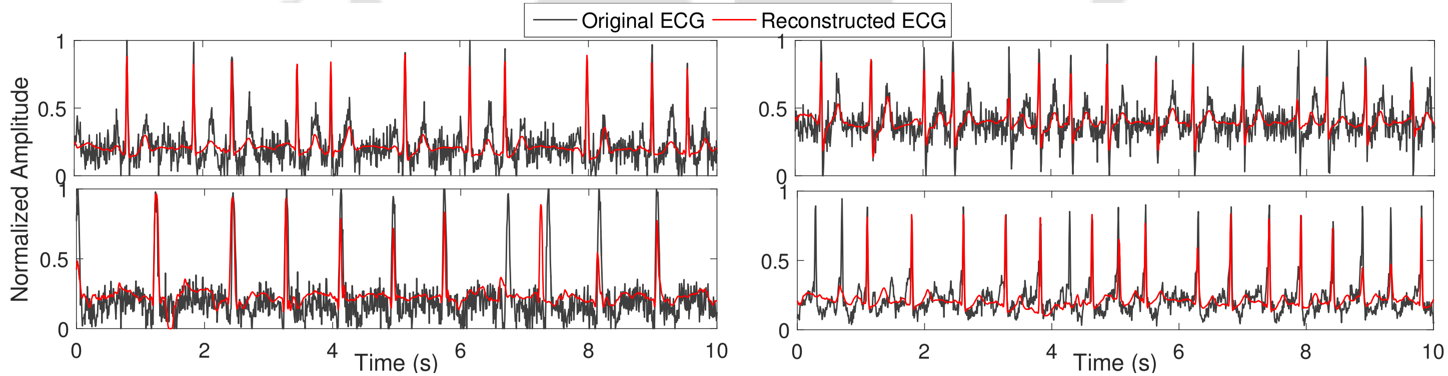


Figure 4.6: Typical AF ECG excepts from the test set with the corresponding reconstructed ECG signals from the CDAE structure of the MT-DCNN model.

4.4.4 Visualization of Reconstructed ECG Signals

Figure 4.6 illustrates typical original noisy ECG signals with their corresponding reconstructed ECG signals from the CDAE structure of the proposed MT-DCNN model. As can be seen, the compact representation vector of the encoder structure can enhance the noisy ECG signals while preserving required diagnostic information related to AF detection, such as irregular rhythm and f-waves. Although the QRS-amplitude has some discrepancies in the reconstructed ECG, still the adequate rhythm information (RR-intervals) with the f-waves information aid in the robust AF detection performance.

4.4.5 Stroke Risk-Stratification in AF

In clinical practice, identifying AF patients at high-risk for stroke is crucial for the timely initiation of life-saving stroke prevention therapies. Moreover, identifying low-risk AF patients is also equally crucial to potentially avoid unnecessary treatments; as the use of oral anticoagulants in low-risk patients carries a risk of bleeding, leading to adverse clinical outcomes [158]. Thus, in this study, following the recent research guidelines [135, 136, 156], a daily AF burden of greater than 4% is used to screen AF patients who are at substantially high-risk and relatively low-risk for stroke, as shown in Figure 4.7 (a) and (b). As can be seen from the distribution of AF burden for the low- and high-risk AF patients, the AF is not a binary entity but rather a spectrum of arterial disease severity that can be further risk-stratified using AF burden measure for its appropriate management.

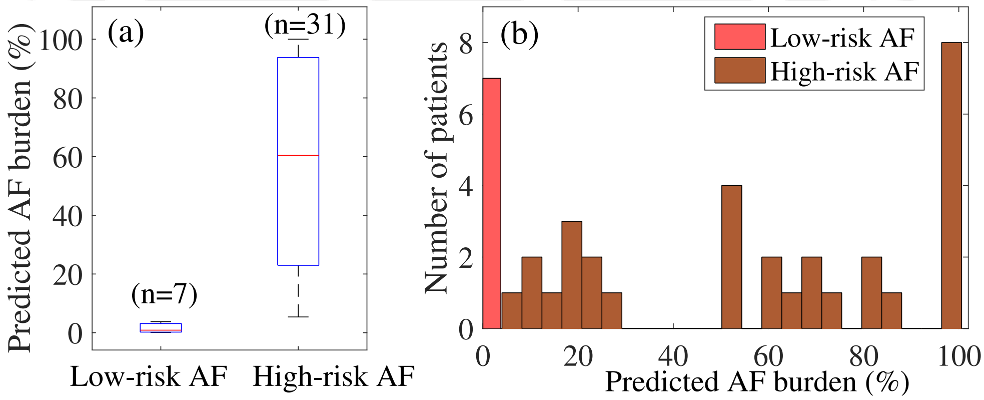


Figure 4.7: (a) Boxplot of predicted AF burden for the AF individuals (n=38) of overall test set with low- and high-risk groups. (b) Histogram of predicted AF burden for the same groups to get further insights about its distribution.

4.4.6 Significance of AF Burden based Evaluation for Improving Diagnosis

The assessment of AF burden estimated from long-term recordings enables the identification of new AF patients who would be missed by the traditional snapshot ECG evaluation. Specifically, patients suffering

Table 4.7: Comparison of AF diagnosis performance for the traditional snapshot ECG evaluation and the proposed AF burden evaluation. TP: true positive, FN: False negative, FP: false positive, and TN: true negative.

Evaluation type	TP	FN	FP	TN	Se(%)	Sp(%)	Acc(%)	F1-score(%)
Snapshot ECG based	**17**	21	2	23	**44.7**	92.0	63.4	60.3
AF burden based	**31**	7	3	22	**81.5**	88.0	84.1	84.6

from paroxysmal AF have high chances of being undiagnosed as they present with sporadic events. We set a threshold of 4$\%$ on the estimated AF burden and performed a binary classification on the overall test dataset (Table 4.7). It resulted in a Se of 81.5$\%$ and Sp of 88.0$\%$. In comparison, using the diagnosis ability of MT-DCNN on a randomly selected ECG excerpt of long-term recording, we report a Se of 44.7$\%$ and Sp of 92.0$\%$. A total of 14 out of 38 additional AF patients (36.8$\%$) are identified using the AF burden-based diagnosis over the snapshot ECG evaluation (Table 4.7). It further emphasizes the importance of using long-term recordings to perform an improved AF diagnosis.

Table 4.8: Comparison of proposed MT-DCNN model with the recent DL-based models in terms of absolute AF burden estimation error ($E_{AF}(\%)$) performance on the LTAFDB test set and three independent validation datasets (AFDB, NSRDB, and LTNSRDB).

Method	LTAF (test)		AFDB		NSRDB		LTNSRDB	
	Raw input	10 dB SNR	Raw input	10 dB SNR	Raw input	10 dB SNR	Raw input	10 dB SNR
ArNet [156]	6.2	10.5	3.3	19.4	5.7	11.8	24.6	61.2
VGGNet [115]	5.6	9.7	3.9	19.8	5.5	14.3	27.1	63.6
1D-CNN [90]	4.9	8.7	6.1	14.7	6.8	13.2	11.7	19.3
MS-CNN [118]	4.3	8.2	5.4	10.9	7.1	13.9	9.5	16.1
Proposed	2.8	5.6	1.7	6.3	3.7	7.1	6.9	13.4

4.4.7 Comparison with Recent DL-based Approaches

Most of the existing DL-based approaches (except [156]) focus on detecting the presence or absence of AF and do not analyze the AF burden using long-term ambulatory ECG recordings. Therefore, we benchmarked the AF burden estimation performance of four representative baseline AF detectors on four PhysioNet datasets and compared them with the MT-DCNN model (Table 4.5 and Figure 4.5). In this section, three recent DL-based AF diagnosis models named VGGNet [115], 1D-CNN [90], and multi-scale CNN (MS-CNN) [118] are adopted for the AF burden estimation from long-term recordings and compared with the proposed MT-DCNN model. In addition, the only existing AF burden estimation model, i.e., ArNet [156], is also considered for the performance comparison. The ArNet [156] and VGGNet [115] take RR-interval and Poincare plot of RR-intervals as input, respectively. On the other hand, 1D-CNN [90], MS-CNN [118], and MT-DCNN models consider ECG excerpts as input. Table 4.8 compares the proposed model and four existing DL-based methods [90, 115, 118, 156] in terms of AF burden estimation error on the LTAF test set and three independent validation datasets. As can be observed, the proposed model outperforms the other four methods at raw ECG and ECG signals corrupted with a 10 dB SNR noise level. Specifically, the models [115, 156] depend on only RR-interval data show more E_{AF} error due to not exploiting P-wave information, which helps discriminate other rhythms with irregular RR-intervals from the AF rhythm. Although models [90, 118] employed complete ECG segment information (atrial and ventricular), their generalization performance on independent validation datasets is limited due to their poor feature representation ability. Compared to existing methods, the proposed multi-task learning model provides lesser E_{AF} error ($p<0.05$) due to its effectiveness in learning better generalizable features. In addition, the proposed MT-DCNN model also shows impressive AF burden estimation performance (lesser E_{AF} error) even at 10 dB SNR over the existing methods, which demonstrates the model's ability to learn robust and generalizable feature representations.

4.5 Summary

The accurate AF burden (the percentage of the time patient is in AF rhythm) estimation from long-term ambulatory ECG recordings has the potential to improve the diagnosis and stroke risk assessment in AF patients and may help guide appropriate personalized stroke prevention strategies. To this end, in this chapter, we demonstrated the feasibility of accurately estimating the AF burden from long-term ambulatory ECG recordings using a new multi-task deep learning model (MT-DCNN). With the innovative incorporation of ECG sequence reconstruction as an auxiliary task for the primary task of detecting AF, the MT-DCNN learned robust feature representations, thereby aiding accurate AF burden estimation from the long-term recordings. Extensive experiments on four diverse Holter ECG datasets demonstrated that the MT-DCNN model outperformed benchmarked feature engineering and other deep learning approaches. Specifically, the proposed model demonstrated robust AF burden estimation in the presence of frequent ectopic beats and different noise levels over the baseline methods. In addition, the analysis of daily AF burden from the long-term recordings showed improved diagnosis and stroke risk assessment for AF over traditional snapshot ECG evaluation. Combining the proposed MT-DCNN model with the recent continuous cardiac health monitoring devices opens a new paradigm of personalized AF diagnostics and management in the context of remote healthcare applications.

5

Multiple Cardiac Disorders Classification from 12-Lead ECG using Multi-Scale Deep Ensemble Approaches

Contents

In the previous chapters, we presented the proposed automated approaches for disease-specific (MI, CHF, and AF) analysis of 12-lead or single-ECG signals. However, in real-world clinical settings, patients can present with any life-threatening cardiac disorder, including MI, CD (left and right BBB), and HYP [13]. Each cardiac ailment has a different pathological origin, i.e., occlusion in the coronary artery causes MI, electrical conduction system dysfunction causes conduction disturbances (BBBs), and abnormal thickening of the heart muscles causes HYP. Therefore, the automated diagnosis systems should have the ability to analyze 12-lead ECG signals of multiple cardiac abnormalities simultaneously and correctly identify the underlying cardiac ailment/ailments for their appropriate management [13, 39]. The reliable automated analysis of multiple cardiac disorders is often challenging due to their subtle and similar pathological ECG manifestations appearing in different shapes and morphologies (multi-scale) based on the abnormality. For example, the appearance of ST-elevation, T-wave peaking and inversion, and pathological Q-waves at the leads facing the injured cardiac muscle characterize the MI disease. The broad and bizarre QRS-complexes, deep and broad S-waves with ST-depressions, and T-inversions in the right-sided (V1/V2) and lateral (I/aVL/V5/V6) ECG leads are often associated with the conduction disturbances (left and right BBBs). Similarly, the increased R-wave amplitudes with ST-depression and T-inversions in the lateral (I/aVL/V5/V6) ECG leads and deep S-waves in the right-sided (V1/V2) leads help to diagnose HYP. It is worth emphasizing that BBBs and HYP disorders are commonly known as ”mimicking MI” diseases owing to their MI-like pathological ECG features [120]. Thus, it is difficult to classify these disorders reliably, considering their similar and multi-scaled pathological ECG morphologies.

Most existing methods [124–129] have employed deep CNN/RNN-based single classifier approaches to classify cardiac disorders. These approaches may not adequately represent the highly variable and multi-scaled ECG manifestations associated with multiple cardiac disorders. For example, although CNN-based models can learn robust features from the ECG signals faster, they fail to capture the temporal dynamics. On the other hand, RNNs can model the temporal dynamics; however, the RNN's sequential processing of input ECG significantly slows down the feature extraction process. The recent developments in the DL-based approaches have led to the design of temporal CNN (TCNN) layers. These layers combine the merits of both CNNs and RNNs and alleviate their limitations using dilated and causal convolutional filters. Specifically, the recent studies [159] demonstrate that the TCNN layers with dilated and causal convolutions outperform the RNNs on various sequence modeling tasks by handling the temporal dependencies effectively. Thus, it is expected that employing TCNNs to model the ECG sequence data can aid in the efficient learning of discriminative features.

In practice, clinicians often seek a second or more opinion for patients with confusing ECG biomarkers before making the final diagnosis. Thus, the weighting of suggestions from several experts can generate a reliable final decision. For example, the suggestion by the specialist cardiologist could get more weight than that of a general physician. Existing methods [76, 119] for multi-class ECG classification used the average-ensemble technique, where the decisions from multiple independently trained DCNN classifiers are aggregated via averaging to obtain the final diagnosis decision (see Figure 5.1 (a)). It is argued that the ensemble methods effectively handle the disease variabilities by combining diagnosis decisions from multiple deep CNN classifiers. However, the average ensemble methods do not consider the interaction among base classifiers during learning. It may limit them from learning specialized discriminative features, thereby degrading the final classification performance. Moreover, these methods assign equal weights to all the base classifiers during decision fusion, irrespective of their competencies. This fusion technique is not data-adaptive and, thus, vulnerable to the weaker deep CNN classifier in the ensemble.

To address the limitations of the average ensemble methods and incorporate the committee of experts-based clinical diagnosis, we explore using a Mixture-of-Experts (MoEs) approach to analyze multiple cardiac ailments. The MoEs is a traditional combining classifiers approach that works on the divide-and-conquer principle. It is often used to tackle a complex classification problem by dividing it among multiple experts whose decisions are combined to yield a final diagnosis. Therefore, inspired by the efficacy of MoEs in medical applications [138, 160], in this chapter, we present a new multi-scale deep temporal convolutional neural network ensemble (MS-DTCE) model that extracts and combines diverse multi-scaled ECG features for effective multi-class ECG classification (see Figure 5.1 (b)). The model consists of several temporal CNN-based expert classifiers designed using dilated and causal convolutional filters with different receptive fields to extract the temporal dynamics of scale-specific pathological ECG characteristics and generate local predictions. A temporal CNN-based gating network is designed to aggregate the local predictions of experts based on their competencies to generate reliable final diagnosis decisions. Moreover, an objective function is formulated to facilitate the interaction and optimal diversity between the experts, enabling them to learn specialized discriminative features for improved classification. This chapter also presents a similar deep temporal CNN ensemble framework for efficient multi-label ECG classification.

The rest of the chapter is organized as follows. Section 5.1 presents the proposed MS-DTCE framework for the multi-class ECG classification. Section 5.2 presents the experimental results and discussion of the MS-DTCE method. The proposed deep temporal CNN ensemble method for multi-label ECG classification and its experimental evaluation are discussed in sections 5.3 and 5.4, respectively. The summary of the chapter is presented in section 5.5.

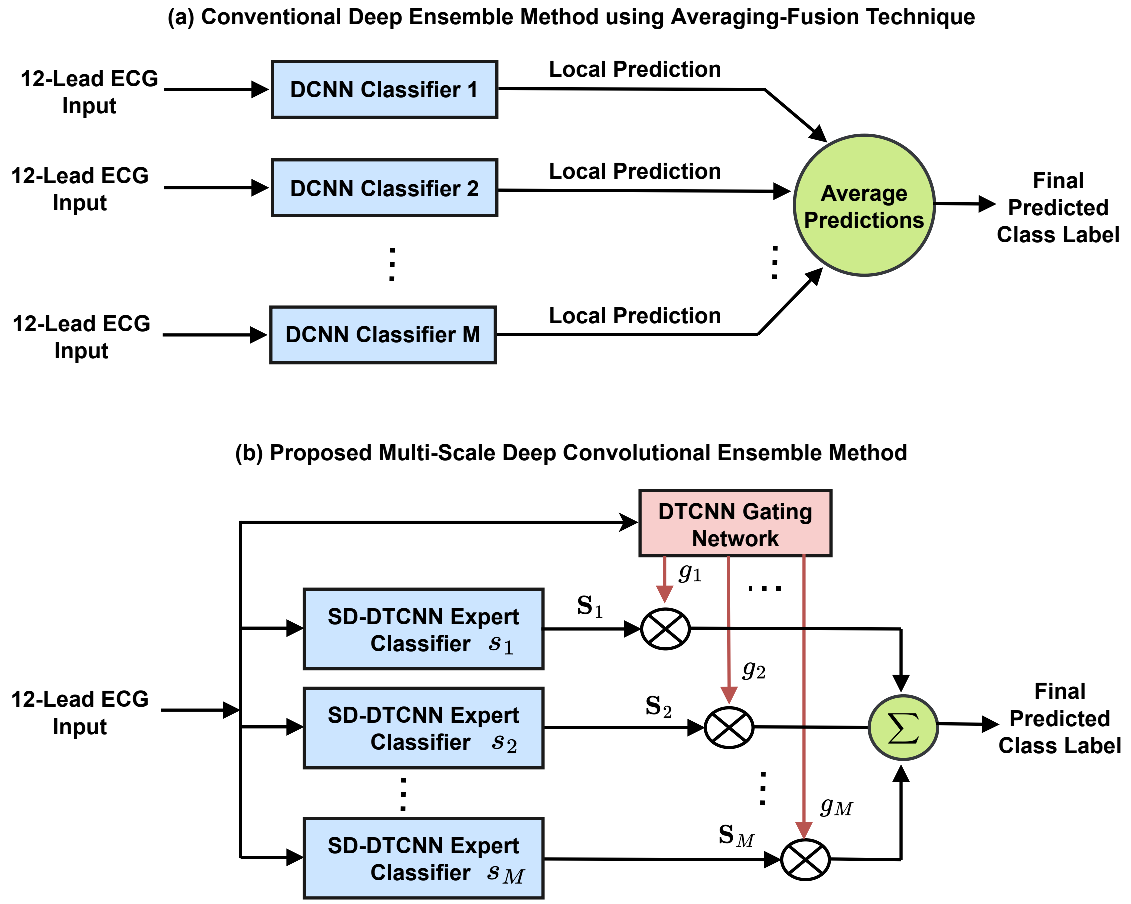


Figure 5.1: (a) Conventional deep ensemble method using averaging-fusion technique. (b) Proposed multi-scale deep temporal convolutional neural network ensemble (MS-DTCE) method.

5.1 Multi-Scale Deep Temporal CNN Ensemble for Multi-Class ECG Classification

This section presents the proposed MS-DTCE method for classifying multiple acute cardiac disorders (MI, CD, HYP, and STTC) and HCs from the 12-lead ECG signals. The block diagram of the method is shown in Figure 5.1 (b). As can be seen, the MS-DTCE is composed of three components: the scale-dependent deep temporal CNN (SD-DTCNN) expert classifiers, the deep temporal CNN (DTCNN) gating network, and finally, the ensemble fusion strategy. The details of each of the components are described as follows.

5.1.1 Scale-Dependent Deep Temporal CNN Expert Classifiers

Pathological ECG characteristics appear in different shapes and morphology across multiple cardiac abnormalities. Such multi-scale characteristics and their temporal dynamics are encoded using multiple SD-DTCNN experts to generate the local predictions. Each expert operates in different receptive fields to effectively capture the multi-scaled disease variabilities. To design SD-DTCNNs with varied receptive fields, we adopted TCNN layers with dilated and causal convolutional filters of different dilation rates. In recent studies [159], the TCNNs with dilated and causal convolutional filters outperform RNNs by effectively modeling the long-term temporal dynamics with the reduced computational burden. Mathematically, for a input sequence $\mathbf{x} \in \mathbb{R}^T$ and a convolutional kernel $\mathbf{w} \in \mathbb{R}^K$, the 1D-dilated causal convolution operation $\mathbf{F}$ on element n of the sequence is defined as:

$$\mathbf{F}[n] = \sum_{i=0}^{K-1} \mathbf{w}[i] \cdot \mathbf{x}[n - d \cdot i] \tag{5.1}$$

where d is the dilation factor, T is the sequence length, K is the filter size and $(n - d \cdot i)$ accounts for the direction of the past samples. Eq. 5.1 can be interpreted as the convolution of input $\mathbf{x}$ and a filter $\mathbf{w}$ with a dilation factor d. The dilated filter is obtained by introducing holes between the kernel elements of $\mathbf{w}$ based on d. As can be seen from Figure 5.2 (a)-(d), the incorporation of holes between the kernel elements increases the effective filter size, thereby increasing the receptive fields. The effective kernel size for a given filter size K and dilation rate d is computed as $(d(K-1)+1)$. In summary, the spacing between the kernel elements helps achieve varied receptive fields in a parameter efficient manner over the traditional convolutions. In addition, the causal nature of filters effectively processes the time-series ECG data.

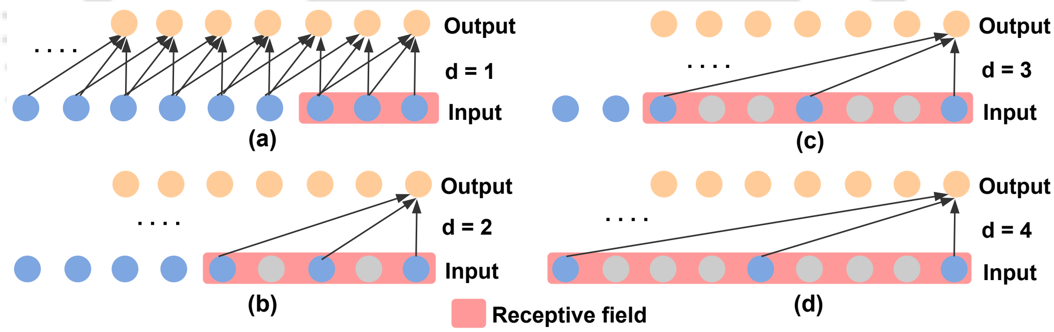


Figure 5.2: A typical visualization of dilated and causal convolution filter with different dilation factors that represent varied receptive field sizes. (a), (b), (c), and (d) show the convolution of input with the filter of size 3 and dilation factor d = 1, 2, 3, and 4, respectively.

Let us denote the 12-lead ECG as $\{\mathbf{X} \in \mathbb{R}^{12 \times T} | \mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^l, ..., \mathbf{x}^{12}]^{\mathsf{T}}\}$. Here, T is the transpose operator and $\mathbf{x}^l \in \mathbb{R}^T$ is the l^{th} ECG-lead with length T. The $\mathbf{X}$ is provided as input to multiple SD-DTCNN expert classifiers $s_d, d \in \{1, 2, ..., M\}$ to obtain the local prediction vectors (see Figure 5.1 (b)). Here, s_d represents SD-DTCNN expert with dilation rate d and M denote the number of experts. Each s_d contains a lead-specific TCNN (LS-TCNN) followed by intra- and inter-lead attention modules (see Figure 5.3). These modules exploit the diversity and integrity aspects of the 12-lead ECG and enhance its feature representation. At first, each-lead of the 12-lead ECG is fed to the LS-TCNN to extract lead-specific features. Figure 5.4 shows the configuration details of LS-TCNN consisting of a stack of dilated causal

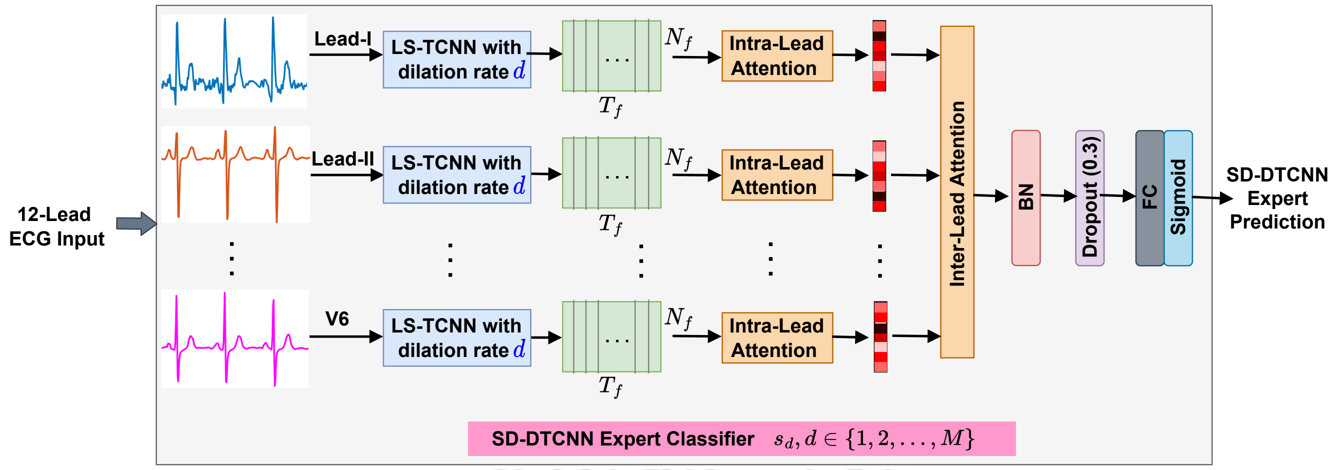


Figure 5.3: Network architecture of the proposed SD-DTCNN expert classifier. For single-lead ECG input, the inter-lead attention module will be discarded.

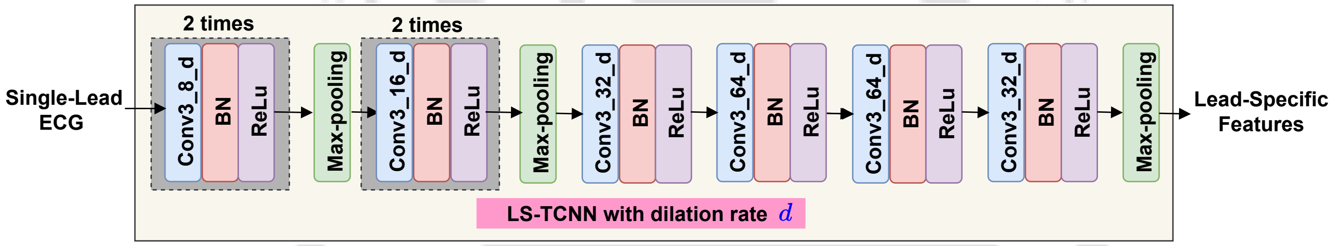


Figure 5.4: Architecture of the LS-TCNN network at dilation factor d. The convolutional layer is represented as "Conv(kernel size)_(number of kernels)_(dilation factor)" and the max-pooling window size is set to 2 with stride 2.

TCNN layers with dilation factor d. The output feature maps of the LS-TCNN can be given as:

$$\mathbf{F}_l = f_l^d(\mathbf{x}^l, \theta_{lead}) \tag{5.2}$$

where f_l^d represents LS-CNN function with dilation rate d and parameters θ_{lead} and it maps l^{th}-lead ECG $\mathbf{x}^l$ to a set of output feature maps $\mathbf{F}_l \in \mathbb{R}^{N_f \times T_f}$. Here, N_f denotes the number of feature maps and T_f denotes the feature map length. Next, the intra-lead attention module generates a temporal attention map $\boldsymbol{\alpha}_l^{Tatt} \in \mathbb{R}^{T_f}$ to summarize the clinically informative temporal information of the LS-TCNN feature maps and produce a within-lead attentive representation (WLAR) vector given as:

$$\mathbf{a}_l = \mathbf{F}_l \boldsymbol{\alpha}_l^{Tatt} \tag{5.3}$$

where $\mathbf{a}_l \in \mathbb{R}^{N_f}, l \in \{1, 2, ..., 12\}$ denotes the WLAR vector for the l^{th} ECG-lead. The LS-TCNN and the intra-lead attention module parameters are shared across the 12-leads to reduce the number of trainable parameters and constrain the network to learn generalizable features. Twelve WLAR vectors from the 12-lead ECG form an attentive matrix $\mathbf{A}_d \in \mathbb{R}^{N_f \times 12}$, and it is fed to the inter-lead attention module. This module generates a lead-attention map $\beta_d^{Latt} \in \mathbb{R}^{12}$ to effectively aggregate the across-lead salient ECG information and generate an across-lead attentive representation (ALAR) vector given as:

$$\mathbf{a}^d = \mathbf{A}_d \beta_d^{Latt} \tag{5.4}$$

where $\mathbf{a}^d \in \mathbb{R}^{N_f}$ denote the high-level ALAR vector and it is fed to the fully-connected (FC) output layer with Sigmoid activation to generate the local prediction vector given as:

$$\mathbf{S}_d = Sigmoid(\mathbf{W}_o \mathbf{a}^d + \mathbf{b}_o) \tag{5.5}$$

where $\mathbf{S}_d \in \mathbb{R}^C$ denotes the local prediction vector from the expert $s_d, d \in \{1, 2, ..., M\}$, M is the number of experts, C is the number of output categories, $\mathbf{W}_o \in \mathbb{R}^{C \times N_f}$ and $\mathbf{b}_o \in \mathbb{R}^C$ are the weights and biases of the output layer, respectively. It should be noted that we adapted the intra- and inter-lead attention mechanism formulations from Chapter 2 to generate the temporal attention map ($\boldsymbol{\alpha}_l^{Tatt}$) and the lead-attention map (β_d^{Latt}), respectively, in the SD-DTCNN experts.

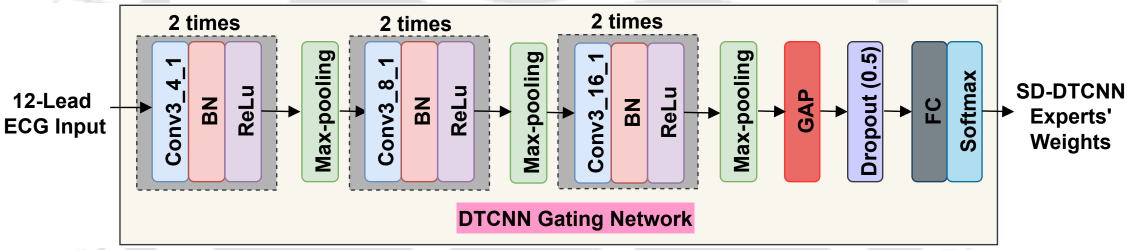


Figure 5.5: Architecture of the DTCNN gating network. The convolutional layer is represented as "Conv(kernel size)_(number of kernels)_(dilation factor)" and the max-pooling window size is set to 2 with stride 2.

5.1.2 Deep Temporal CNN Gating Network

To aggregate the experts' local prediction vectors in a data-adaptive manner, we employ a DTCNN gating network. As shown in Figure 5.5, the gating network computes the weights of importance for the experts according to their competencies in classifying the input ECG signals. A global average pooling (GAP) is applied to summarize the convolutional feature maps by squeezing the temporal dimension (see Appendix A). Finally, a FC output layer with Softmax activation is used to generate the fusion weights as:

$$\mathbf{g} = Softmax(f_g(\mathbf{X}, \theta_g)) \tag{5.6}$$

where $\mathbf{g} \in \mathbb{R}^M$ is the fusion weights vector for the M experts and f_g represents the DTCNN gating function (excluding output layer) with parameters θ_g.

5.1.3 Ensemble Fusion Strategy

The proposed method generates multiple local prediction vectors $\mathbf{S}_d, d \in \{1, 2, ..., M\}$ from the M experts and their weights of importance $\mathbf{g} \in \mathbb{R}^M$ from the gating network. The final prediction vector of the MS-DTCE model for the input $\mathbf{X}$ is denoted as $\mathbf{F}_{MS-DTCE}(\mathbf{X})$ and is given by:

$$\mathbf{F}_{MS-DTCE}(\mathbf{X}) = p(c|\mathbf{X}) = \sum_{d=1}^{M} g_d(\mathbf{X}) \cdot \mathbf{S}_d(\mathbf{X}) \tag{5.7}$$

where $p(c|\mathbf{X})$ is the probability of input $\mathbf{X}$ belongs to class $c, c \in \{1, 2, ..., C\}$ and g_d denote the weight of importance for d^{th} expert prediction vector. The final predicted output class label for the input ECG $\mathbf{X}$ is obtained as:

$$class(\mathbf{X}) = \underset{c=1,2,...,C}{\arg\max}\; p(c|\mathbf{X}) \tag{5.8}$$

5.1.4 Training of the MS-DTCE Architecture

The optimal diversity among the individual experts is crucial for ensemble methods to achieve improved classification performance [161]. The conventional MoEs method can produce diverse experts using the error function based on the negative log probability of generating the desired output vector under the mixture of Gaussian models [137]. However, the MoEs error function does not include any control parameter to maintain the optimal diversity among the experts. Inspired by the negative correlation learning methods [161], we have incorporated an explicitly controllable cross-correlation penalty term to the MoEs error function. The formulation of proposed MS-DTCE error objective function for the input ECG $\mathbf{X}$ ($\mathbf{E}_{MS-DTCE}(\mathbf{X})$) is given by:

$$\mathbf{E}_{MS-DTCE}(\mathbf{X}) = -log\Bigg(\sum_{d=1}^{M} g_d(\mathbf{X}) exp\Big(-\frac{1}{2}\|\mathbf{y} - \mathbf{S}_d(\mathbf{X})\|_2^2 + \lambda r_d\Big)\Bigg), \lambda \geq 0 \tag{5.9}$$

where $\mathbf{y}$ denotes the one-hot encoded target output vector of input ECG $\mathbf{X}$. $\mathbf{S}_d$, g_d and r_d represents the local prediction vector, the weight of importance and the correlation penalty term of d^{th} SD-DTCNN expert classifier, respectively. M denotes the number of expert classifiers in the ensemble. As can be observed in Eq. 5.9, the proposed error objective function consists of two parts: (i) negative log probability of generating the desired output vector under the mixture of Gaussian models, minimizing this term enables improved overall classification performance, and (ii) cross-correlation penalty term, minimizing this term enables optimal diversity among the experts. The parameter λ denotes the strength of the penalty, which controls the trade-off between overall accuracy and experts' diversity of the proposed ensemble approach. The correlation penalty term r_d is computed as follows:

$$r_d = \big(\mathbf{S}_d(\mathbf{X}) - \mathbf{F}_{MS-DTCE}(\mathbf{X})\big) \sum_{j=1;j\neq d}^{M} \big(\mathbf{S}_j(\mathbf{X}) - \mathbf{F}_{MS-DTCE}(\mathbf{X})\big)^{\mathsf{T}} \tag{5.10}$$

where $\mathbf{F}_{MS-DTCE}(\mathbf{X})$ denote the final prediction vector of the MS-DTCE model for the input **X** and is computed by Eq. 5.7. It can be seen from Eq. 5.9 that when $\lambda = 0$, there is no correlation penalty term in the MS-DTCE error function. From Eq. 5.10, the importance of the penalty term is to negatively correlate each experts' error with the other experts' errors, thereby regularizing the experts to learn specialized disease characteristics. It is worth emphasizing that the proposed error objective (Eq. 5.9) is a function of of DTCNN gating network output weights $g_d, d \in \{1, 2, ..., M\}$ and the experts' local prediction vectors $\mathbf{S}_d, d \in \{1, 2, ..., M\}$. Thus, the error function enables simultaneous and interactive learning of all the experts and the gating network in an end-to-end manner for improved overall classification performance.

5.2 Results and Discussion for the Multi-Class ECG Classification

This section presents the details of the clinical ECG datasets, the signal preprocessing, and the evaluation metrics employed. This section also presents the details of the network optimization and the experimental results of the proposed and baseline methods, followed by the analysis of the proposed MS-DTCE model.

5.2.1 Clinical ECG Database

The proposed deep ensemble classification method is evaluated on two publicly available databases from the PhysioNet repository [148], i.e., the PTBXL-2020 12-lead ECG database and the PhysioNet/CinC-2017 single-lead ECG database. The specific details of these databases are given below.

5.2.1.1 PTBXL-2020 Database

The PTBXL-2020 is the largest publicly available clinical 12-lead ECG dataset [162] comprising 21837 records of 10 s duration from 18885 patients, recorded by the Schiller AG ECG devices with a sampling rate of 500 Hz. The database consists of four disease categories, including conduction disturbance (CD), hypertrophy (HYP), myocardial infarction (MI), ST/T changes (STTC), and a normal ECG (NORM) category. In this study, we chose ECG records having single-label annotations for the multi-class ECG classification and excluded multi-label ECG records from the analysis. The authors of the dataset provided the training, validation, and test data splits that we employed for the method evaluation. Table 5.1 shows the details of the selected 12-lead ECG dataset.

Table 5.1: Data details for the PTBXL-2020 and the PhysioNet/CinC-2017 datasets

PTBXL-2020 (12-lead ECG dataset)					**CinC-training2017** (single-lead ECG dataset)	
Class	**# Records**	**Training**	**Validation**	**Test**	**Class**	**# Records**
NORM	8981	7180	899	902	N	5154
CD	1691	1338	170	183	AF	771
HYP	538	418	64	56	OTH	2557
MI	2513	2026	232	255	Noisy	46
STTC	2380	1886	252	242	–	–
Total	16103	12848	1617	1638	**Total**	8528

#: Number of.

5.2.1.2 PhysioNet/CinC-2017 Dataset

This dataset [163] comprises 8528 single-lead ECG records ranging from 9 to 61 s, recorded from Alivecor's single-channel ECG hand-held device and sampled at 300 Hz. In this study, the three-class types, including normal (N), atrial fibrillation (AF), and other (OTH) rhythms, are used for the analysis. A five-fold cross-validation technique at the patient level is used to evaluate the average performance of the method on this dataset. Table 5.1 shows the details of the CinC-2017 dataset.

5.2.2 Data Preprocessing

5.2.2.1 Downsampling and Baseline Artifact Removal

All the ECG signals are downsampled to 100 Hz to reduce the computational burden on the proposed model. The practical ECG signals are often corrupted with the baseline artifact, which significantly changes the ECG amplitudes. Thus, to remove this artifact, ECG signals are passed through a Butterworth low-pass filter with a cut-off frequency of 0.5 Hz; and the filter output is later subtracted from the original signal.

5.2.2.2 Cropping, Padding and Data Normalization

The input ECG records can have varying lengths. Therefore, we choose a fixed length of 20 s for the CinC-2017 database as most of its records last at least 20 s [118]. The longer ECG records are cropped, and shorter records are padded with zeros to maintain a 20 s duration. It is to be noted that no cropping and padding is performed on the ECG records of the PTBXL-2020 dataset as all the signals are of the same length, 10 s. Moreover, the ECG recordings amplitudes vary among different subjects and databases. Thus, each ECG record is z-score normalized to maintain zero mean and unit variance.

5.2.2.3 Data Augmentation

The PTBXL-2020 and the CinC-2020 datasets show class imbalance (Table 5.1) with more normal class ECG records over other classes. This imbalance in sample distribution has a negative impact on training DL models as the model would bias towards the dominating classes. To mitigate this problem, two data augmentation techniques, i.e., (i) adding random additive white Gaussian noise and (ii) small amplitude and duration perturbations, are employed for the small sample disease classes [126]. It is to be noted that the data augmentation is performed only on the training dataset, and no such augmentation is performed on the validation and test sets.

5.2.3 Performance Evaluation Metrics

To evaluate the classification performance, we have used the class-wise F1-score and the macro-average F1-score, macro-average precision, macro-average recall, accuracy (Acc) and Kappa value metrics (see Appendix A). To evaluate the disagreement between the experts, we have used the distance-based disagreement (DbD) measure [164]. It represents the average expert's disagreement and is computed from the confusion matrix distances of individual experts. A higher DbD value indicates more disagreement between the experts for class prediction. The proposed method with M experts, a distance metric D^m between expert m and all other experts is computed as:

$$D^m = \frac{1}{C}\sum_{i=1}^{C}\sum_{j=1}^{C}|\mathbf{U}_{i,j}^m - \sum_{k=1;k\neq m}^{M}\mathbf{U}_{i,j}^k| \tag{5.11}$$

where $\mathbf{U}_{i,j}^m$ represents the confusion matrix elements of expert m, $m \in \{1, 2, ..., M\}$ with C classes. Finally, the average DbD measure (DbD $\geq$ 0) can be computed as follows:

$$DbD = \frac{1}{M}\sum_{m=1}^{M} D^m \tag{5.12}$$

5.2.4 Network Parameters and Training Settings

The preprocessed ECG signals are fed to the proposed MS-DTCE framework to obtain the cardiac disorder class label. The network architecture details of the method are provided in Figure 5.3, Figure 5.4 and Figure 5.5. In this study, the SD-DTCNN experts ($s_d, d \in \{1, 2, ..., M\}$) combination and the strength of correlation penalty (λ) are set through experimentation. The SD-DTCNN combination $s_1 - s_2 - s_4$ at $\lambda = 0.2$ provides best $F1$-score for the PTBXL-2020 dataset. Similarly, the experts combination $s_1 - s_2 - s_4$ at $\lambda = 0.3$ provides best $F1$-score for the CinC-2017 dataset (will be discussed in upcoming sections). The

MS-DTCE network is developed in Keras using the TensorFlow backend. An NVIDIA Tesla V100 GPU is used to perform the experiments. All the parameters of the proposed architecture are trained using the MS-DTCE error function (Eq. 5.9) in an end-to-end manner. The model is trained with an Adam optimizer with a learning rate of 10^{-3} on mini-batches of size 32 for 100 epochs. To alleviate overfitting, an early stopping criterion is employed; which stops the training when the validation F1-score does not improve over 20 iterations. The best-performing model is used for the performance evaluation.

5.2.5 Baseline Methods used for Performance Comparison

To verify the effectiveness of the proposed MS-DTCE method, we have compared it with the following DL-based state-of-the-art 12-lead/single-lead ECG classification approaches.

- *Single deep classifier based non-ensemble methods:*
  - *Single DCNN/DRNN based methods:* These methods employ single DCNN/DRNN models for the feature extraction followed by an FC+Softmax layer for the multi-class ECG classification. In this category, the proposed method is compared with DCNN [165] method, fully convolutional neural network (FCNN) method [76], multi-lead residual neural network (ML-ResNet) [166], a hybrid network using CNN and bi-directional recurrent neural networks called MFB-CRNN [78], attention-based time-incremental CNN (ATI-CNN) method [126] and multi-lead diagnostic attention-based DRNN (MLDA-RNN) method [167] for the PTBXL-2020 dataset. Similarly, for the CinC-2017 dataset, VGG12 [168], ResNet50 [153], DCNN [165] method, temporal attention-based CNN (TA-CNN) [169] and sequential LSTM [170] method are used for the comparison.
  - *Concatenation fusion methods:* These methods use several DCNN branches for the multi-scale feature extraction. The output features of all the DCNNs are fused via concatenation followed by an FC+Softmax layer for the cardaic disorders classification. In this category, the MS-DTCE is compared with the multi-scaled fusion CNN (MS-CNN) [118] and the deep multi-scale fusion neural network (DMSFNet) [129].
- *Average ensemble methods:* The averaging ensemble is the commonly used strategy to design deep ensemble methods for arrhythmia classification [76, 119]. In this approach first, several DCNN-based classifiers are trained independently to generate local predictions. Later, these individual predictions are averaged to obtain the final diagnosis decision. For comparison, the average ensemble approach

is designed by averaging the predictions from all the single DCNN/DRNN methods [76, 78, 126, 165–167] and the concatenation fusion methods [118, 129] for the PTBXL-2020 dataset.

It is to be noted that the single DCNN/DRNN methods [76, 78, 126, 165–167] and the concatenation fusion methods [118, 129] were originally proposed for the arrhythmia or MI classification using the ECG signals. As the PTBXL-2020 is a newly released database, the network parameters of the above exiting methods are learned (benchmarked) on the PTBXL-2020 database with the same data processing, splitting, and training iterations as the proposed method for a fair comparison. On the other hand, for the CinC-2017 database, the existing methods' [118, 129, 165, 169] validation results are directly taken from the research articles, as a similar evaluation strategy on the same dataset is adopted in this chapter.

5.2.6 Multi-Class ECG Classification Results

5.2.6.1 Evaluation on the PTBXL-2020 Dataset

Table 5.2 compares the proposed MS-DTCE and the baseline methods in terms of class-wise F1-score (F1N: normal, F1C: CD, F1H: HYP, F1M: MI and F1S: STTC) and the overall measures on the PTBXL-2020 test set. As can be seen, the proposed MS-DTCE outperforms existing methods across all the evaluation metrics. Specifically, the single deep classifier based DCNN/DRNN models [76, 78, 126, 165–167] provide poor classification results. The attention-based ATI-CNN [126] and MLDA-RNN [167] methods provide a slight improvement in the performance compared to the other DCNN/DRNN models by summarizing the informative ECG features. However, these single DCNN models may not adequately represent the multi-scale ECG variabilities; thus, shows limited classification performance. Compared to the best performing single DCNN method (ATI-CNN [126]), the proposed MS-DTCE provides an improvement of nearly 9.41% in precision, 9.09% in recall and 9.82% in F1-score.

The concatenation fusion methods, i.e., the MS-CNN [118] and the DMSFNet [129] show improved performance over the single DCNN/DRNN models by extracting and integrating multi-scale ECG features using a two-stream DCNN models. However, these methods' performance is limited by the sub-optimal DCNN feature maps fusion via concatenation. Compared to the best performing concatenation fusion method DMSFNet [129], the proposed method exhibits impressive results with an improvement of nearly 6.70% in precision, 7.08% in recall and 7.57% in F1-score (Table 5.2). The average ensemble method shows minimal improvement over its base classifier methods [76, 78, 118, 126, 129, 165–167] due to (i) the lack of interaction among the base classifiers during training and (ii) assigning equal weight of importance for the base classifier predictions during the decisions fusion. Compared to the average ensemble approach, the proposed method provides an improvement of nearly 6.33% in precision, 12.61% in recall and 9.65%

Table 5.2: Performance comparison of the proposed and baseline methods on test set of the PTBXL-2020 database.

Method	Configuration	Performance Measures									
		F1N(%)	**F1C**(%)	**F1H**(%)	**F1M**(%)	**F1S**(%)	**Pr**(%)	**Re**(%)	**F1**(%)	**Acc**(%)	**Ka**(%)
Single DCNN/DRNN methods†	DCNN [165]	83.99	71.01	45.96	61.67	68.65	65.80	68.93	66.26	75.15	61.13
	FCNN [76]	85.08	59.04	65.01	63.18	67.80	68.12	70.53	68.02	74.96	61.74
	ML-ResNet [166]	82.59	72.77	56.96	67.13	69.49	66.46	75.87	69.79	75.51	63.68
	MFB-CBRNN [78]	83.89	69.76	36.62	71.01	69.93	65.90	68.06	66.24	76.25	63.12
	ATI-CNN [126]	86.60	78.31	65.12	71.78	71.82	74.84	76.12	74.73	80.40	69.27
	MLDA-RNN [167]	84.73	75.78	47.41	72.38	72.58	70.67	71.24	70.58	78.63	66.23
Concatenation fusion†	MS-CNN [118]	85.68	72.14	78.32	64.62	67.54	71.76	77.58	73.66	77.83	65.79
	DMSFNet [129]	86.84	77.19	82.44	67.58	70.84	77.55	78.13	76.98	80.58	68.98
Average ensemble	[76, 78, 118, 126, 129, 165–167]	86.87	75.07	69.09	69.45	73.33	77.92	72.60	74.90	80.76	68.58
Proposed concatenation fusion	s_1, s_2, s_4	87.56	78.90	68.90	73.02	74.15	74.19	81.17	76.40	81.50	71.40
Proposed average ensemble	s_1, s_2, s_4	87.05	75.64	72.27	69.72	72.85	77.86	74.04	75.51	80.95	68.87
Proposed MS-DTCE ($\lambda = 0.2$)	$\mathbf{s_1 - s_2 - s_4}$	**89.65**	**82.35**	**93.33**	**78.61**	**78.78**	**84.25**	**85.21**	**84.55**	**85.65**	**77.45**

F1N, F1C, F1H, F1M, and F1S: Class-wise F1-score for NORM, CD, HYP, MI, and STTC, respectively.
Pr: Precision, Re: Recall, F1: F1-score, Acc: Accuracy, and Ka: Kappa score.
†: These methods are trained on the recent PTBXL-2020 database for comparison.

in F1-score, which implies the effectiveness of the proposed ensemble method. Similar improvement in performance can be observed for other measures as well.

To verify the effectiveness of the proposed ensemble method, we have adopted the SD-DTCNN experts s_1, s_2 and s_4 to the concatenation fusion and the average ensemble frameworks. The results for the proposed concatenation fusion are obtained by fusing the features (obtained prior to the output layer) from experts through concatenation and provided them to an FC+Softmax layer. For the proposed average ensemble approach, the local predictions from the experts are fused via averaging to obtain the final classification decision. The results for both the approaches are shown in Table 5.2. As discussed earlier, both approaches show limited classification results compared to the proposed MS-DTCE method. In summary, the proposed method ($s_1 - s_2 - s_4$ at $\lambda = 0.2$) exhibits improved classification results with precision of 84.25%, recall of 85.21% and F1-score of 84.55%.

Furthermore, a drop in class-wise F1-score for the disease classes CD, HYP, MI and STTC is observed across all the methods (Table 5.2) because these class types often show subtle discrimination in 12-lead ECG features; thus are difficult to discriminate. The small sample size might have further degraded the performance. However, compared to existing methods, the discrimination between CD, HYP, MI and STTC classes is higher for the proposed MS-DTCE (see class-wise F1-scores from Table 5.2). In particular, over average ensemble method, the proposed MS-DTCE method shows an F1-score improvement of 7.28% for CD class, 24.24% for HYP class, 9.16% for MI class and 5.45% for STTC class. The improved classification results of the proposed method demonstrate its effectiveness in performing complex classification tasks.

Table 5.3: Performance comparison of the proposed and baseline methods on the CinC-2017 dataset (5 folds).

Method	Configuration	Performance Measures					
		F1NR(%)	**F1A**(%)	**F1O**(%)	**Precision**(%)	**Recall**(%)	**F1**(%)
Single DCNN/DRNN methods†	VGG12 [168]	88	75	69	80.0	76.4	77.9
	ResNet50 [153]	89	72	71	79.8	76.1	77.7
	DCNN [165]	89	72	69	79.5	75.2	77.0
	TA-CNN [169]	92	82	71	N/A	N/A	82.3
	LSTM [170]	90	68	71	81.3	73.4	76.2
Concatenation fusion†	MS-CNN [118]	90	77	74	82.6	79.3	80.7
	DMSFNet [129]	92	83	78	85.6	82.9	84.1
Proposed concatenation fusion	s_1, s_2, s_4	93	79	81	83.8	86.0	84.5
Proposed average ensemble	s_1, s_2, s_4	91	80	81	84.1	83.6	84.0
Proposed MS-DTCE ($\lambda = 0.3$)	$\mathbf{s_1 - s_2 - s_4}$	**95**	**86**	**84**	**88.4**	**88.2**	**88.3**

F1NR, F1A, and F1O: Class-wise F1-score for NSR/HC, AF, and OTH, respectively. F1: F1-score.
†: The results of these methods are taken from the research articles [118, 129, 165, 169].

5.2.6.2 *Evaluation on the PhysioNet/CinC-training2017 Dataset*

Table 5.3 compares the proposed and the baseline methods in terms of class-wise F1-score (F1NR: normal rhythm, F1A: AF and F1O: other rhythm) and the overall measures for the CinC-2017 database. Following the literature studies, we have also employed a 5-fold cross-validation to evaluate the proposed method. As can be seen from Table 5.3, the proposed MS-DTCE ($s_1 - s_2 - s_4$ at $\lambda = 0.3$) outperforms existing DL-based methods with an average F1-score of 88.3$\%$. In particular, compared to the single deep classifier-based DCNN/DRNN methods [153, 165, 168–170], the proposed method shows a minimum improvement of nearly 6$\%$ in the F1-score. Over the concatenation fusion methods MS-CNN [118], DMSFNet [129], the proposed MS-DTCE shows 7.6$\%$ and 4.2$\%$ improvement in the F1-score, respectively. The MS-DTCE also outperforms the proposed concatenation fusion and average ensemble frameworks of the SD-DTCNN experts s_1, s_2 and s_4 with a minimum improvement of nearly 3.8$\%$ in F1-score. Furthermore, compared to the conventional average ensemble method, the proposed MS-DTCE exhibits an improvement of nearly 4$\%$, 6$\%$ and 4$\%$ in NSR-, AF-, and other-class F1-score, respectively (Table 5.3). Similar improvements can also be observed for the class-wise F1-scores and the precision and recall measures of the proposed MS-DTCE over baseline methods. These impressive classification results indicate that the proposed method can effectively extract various multi-scale disease-related features and combine them to improve the discrimination capability of the model.

In order to verify the diversity among the individual SD-DTCNN experts, we have computed the DbD measure for the ensemble methods, i.e., the average ensemble method, the proposed method with no

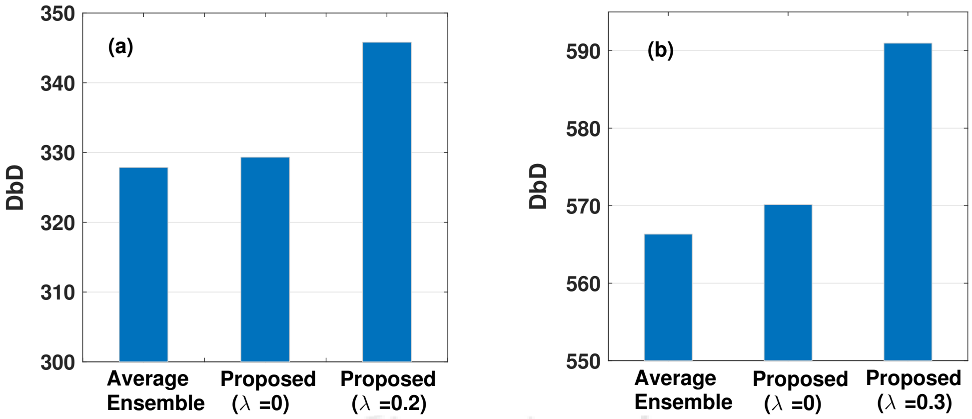


Figure 5.6: Comparing DbD value of the proposed average ensemble and the proposed MS-DTCE at λ=0 and λ=0.2 or 0.3 for (a) the five-class PTBXL-2020 database and (b) the three-class CinC-2017 database, respectively.

correlation penalty in the error function and the proposed method with optimal correlation penalty in the error function. Figure 5.6 (a) and Figure 5.6 (b) show the DbD measure for the above methods across the PTBXL-2020 and the PysioNet/CinC-2017 datasets, respectively. As can be seen, the proposed method at optimal λ shows a higher DbD value for both datasets. This implies that the correlation penalty term in the MS-DTCE error function enables optimal diversity among the experts and helps their ensemble achieve improved classification performance.

5.2.7 Effectiveness of the Proposed MS-DTCE Method

In this section, we analyze the effectiveness of each component of the proposed MT-DTCE method, such as the strength of penalty λ, the multi-scale expert's ensemble, and the attention mechanism on the classification performance. This section also presents the model transparency visualizations and the model complexity regarding the number of trainable parameters.

5.2.7.1 Effect of λ Parameter on the Classification Performance

The parameter λ in the MS-DTCE error function (Eq. 5.9) controls the trade-off between overall classification accuracy and diversity among SD-DTCNN experts. Therefore, the proper selection of λ is crucial for achieving optimal diversity among the experts so that their decision fusion may lead to improved classification performance. Figure 5.7 (a) and Figure 5.7 (b) show the variation of F1-score concerning the λ parameter across various MS-DTCE architectures for the PTBXL-2020 and the CinC-2017 datasets, respectively. As can be seen, with increasing λ from 0, there is a significant improvement in the F1-score for all the MS-DTCE structures. Specifically, at $\lambda = 0.2$ and $\lambda = 0.3$ the proposed $s_1 - s_2 - s_4$ ensemble achieves best classification results for the PTBXL-2020 dataset and the CinC-2017 dataset, respectively.

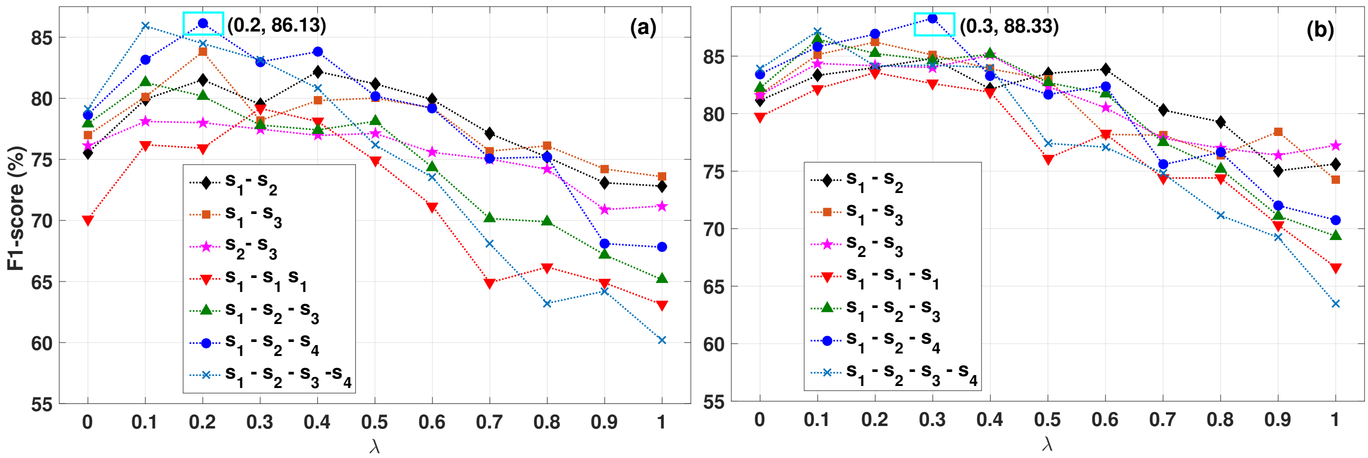


Figure 5.7: Dependency of various MS-DTCE architectures' F1-score on the control parameter λ for (a) the PTBXL-2020 validation dataset and (b) the CinC-2017 dataset.

However, increasing the λ value beyond the optimal degrades the classification performance. This degradation is because, at the higher λ values, the MS-DTCE learning put relatively more emphasis on creating diverse SD-DTCNN experts over classification accuracy. Therefore, it implies that too diverse (beyond optimal) experts in the ensemble may not be effective for making reliable diagnosis decisions.

Table 5.4: Performance comparison of various MS-DTCE architectures.

	PTBXL-2020 Dataset		CinC-2017 Dataset	
Configuration	**Optimal λ**	**F1(%)**	**Optimal λ**	**F1(%)**
s_1	N/A	74.76	N/A	80.5
s_2	N/A	74.26	N/A	81.6
s_3	N/A	72.85	N/A	83.5
s_4	N/A	73.45	N/A	82.3
$s_1 - s_2$	0.4	78.34	0.3	84.8
$s_1 - s_3$	0.2	78.01	0.2	86.2
$s_2 - s_3$	0.1	74.09	0.4	85.1
$s_1 - s_1 - s_1$	0.3	74.62	0.2	83.6
$s_1 - s_2 - s_3$	0.1	80.10	0.1	86.4
$\mathbf{s_1 - s_2 - s_4}$	**0.2**	**84.55**	**0.3**	**88.3**
$s_1 - s_2 - s_3 - s_4$	0.1	84.09	0.1	87.2

N/A: Not applicable.

5.2.7.2 *Significance of Multi-Scale Experts Ensemble*

The classification performance of the proposed method is effectively improved by the ensemble of multiple SD-DTCNN expert classifiers. To substantiate the above claim, we compare the F1-score of various single-scale and multi-scale combinations of the proposed structures for the PTBXL-2020 and the CinC-2017 datasets shown in Table 5.4. It can be observed that the single-scale DTCNNs (s_1, s_2, s_3 and s_4) are not sufficient for the reliable cardiac disease classification. However, the proposed multi-scale MS-DTCE combinations significantly improved the classification performance. Specifically, among various SD-DTCNN expert combinations, the $s_1 - s_2 - s_4$ ensemble at $\lambda = 0.2$ and $\lambda = 0.3$ generate improved F1-score for the PTBXL-2020 and the CinC-2017 datasets, respectively. This verifies that the fusion of decisions from multiple SD-DTCNN experts is crucial for exploiting key multi-scale ECG variabilities, thereby improving the decision-making capability of the proposed MS-DTCE model.

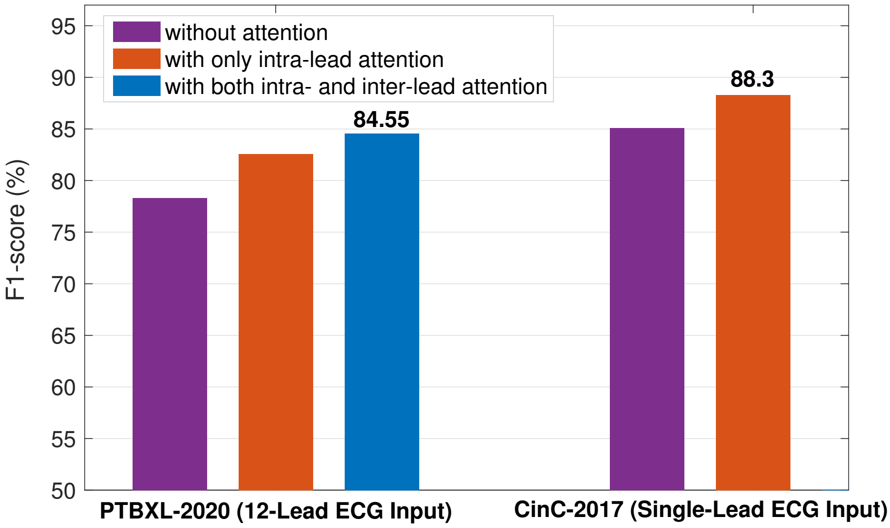


Figure 5.8: F1-score comparison of the MS-DTCE model without attention, with only intra-lead attention and with both intra- and inter-lead attention (valid only for 12-lead ECG input) for the PTBXL-2020 and the CinC-2017 datasets.

5.2.7.3 Influence of Attention Mechanism

The intra- and inter-lead attention modules present in each expert can also affect the classification performance. To verify this, we implement the MS-DTCE model without the attention mechanism by removing the intra- and inter-lead attention modules from the SD-DTCNN experts, i.e., in each expert, the LS-TCNN feature maps are averaged and fed to the output FC+Softmax layer for class prediction. Similarly, for the MS-DTCE model with only an intra-lead attention module, i.e., in each SD-DTCNN expert, an intra-lead attention module is incorporated to summarize the clinically informative temporal information and generate a WLAR vector. These WLAR vectors across the 12-leads are averaged and fed to the output FC+Softmax layer. It should be noted that for the single-lead ECG input from the CinC-2017 dataset, the WLAR vector is directly fed to the output FC+Softmax layer for the classification. Figure 5.8 shows the F1-score comparison of the MS-DTCE model without and with attention modules on the PTBXL-2020 and the CinC-2017 datasets. As can be seen, the MS-DTCE model with attention modules generates improved classification performance by summarizing relevant within and across leads information.

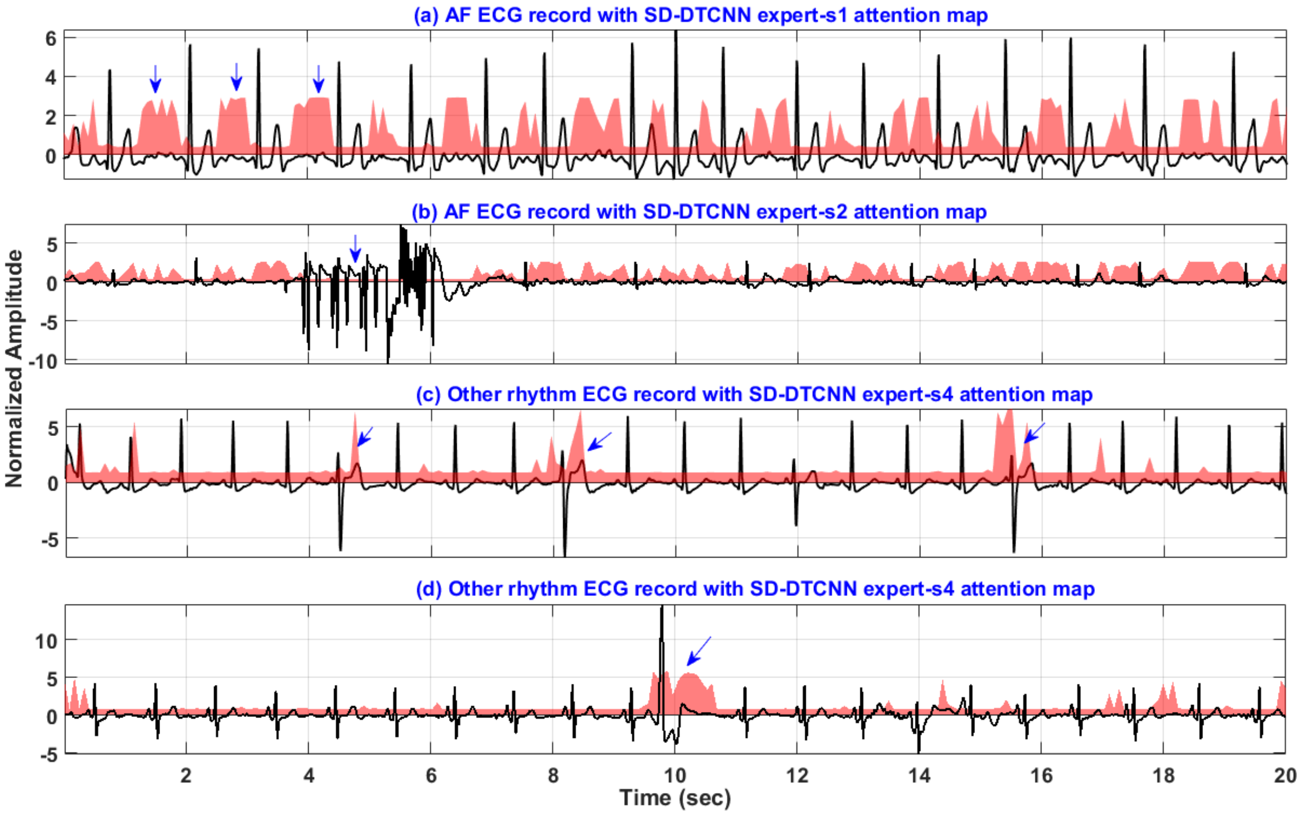


Figure 5.9: Visualization of intra-lead attention weights (red color maps) assigned to the different segments of ECG records (from the CinC-2017 dataset) by the best SD-DTCNN expert in the ensemble. (a)-(b) and (c)-(d) show the AF and other rhythm ECG records with their attention maps, respectively.

5.2.7.4 Model Interpretability

The applicability of deep learning models in the healthcare domain is often limited because these methods lack transparency and are perceived as black boxes by doctors. Recently, the attention-based frameworks have partially succeeded in alleviating this issue [129, 167]. In this direction, we qualitatively evaluate the proposed method to provide a transparent diagnostic basis by visualizing the intra-lead attention maps from the best SD-DTCNN expert in the ensemble. Figure 5.9 shows attention maps for a typical AF (Figure 5.9 (a)-(b)) and other rhythm (Figure 5.9 (c)-(d)) ECG records from the PhysioNet/CinC-2017 dataset. As can be seen, the intra-lead attention module assigned different weights to different ECG-segments of the records. Clinically AF is characterized by the absence of P-waves or the presence of fluttery P-waves in the ECG. As observed from Figure 5.9 (a)-(b), the attention maps indeed focus on these abnormal P-waves (just before R-peak) over QRS- and T-waves during the diagnosis. In particular, Figure 5.9 (b) shows the AF ECG record corrupted with transient noise artifact in between 4-6 s. Interestingly, the model assigned smaller attention weights to the noisy region and larger attention weights to the abnormal P-waves

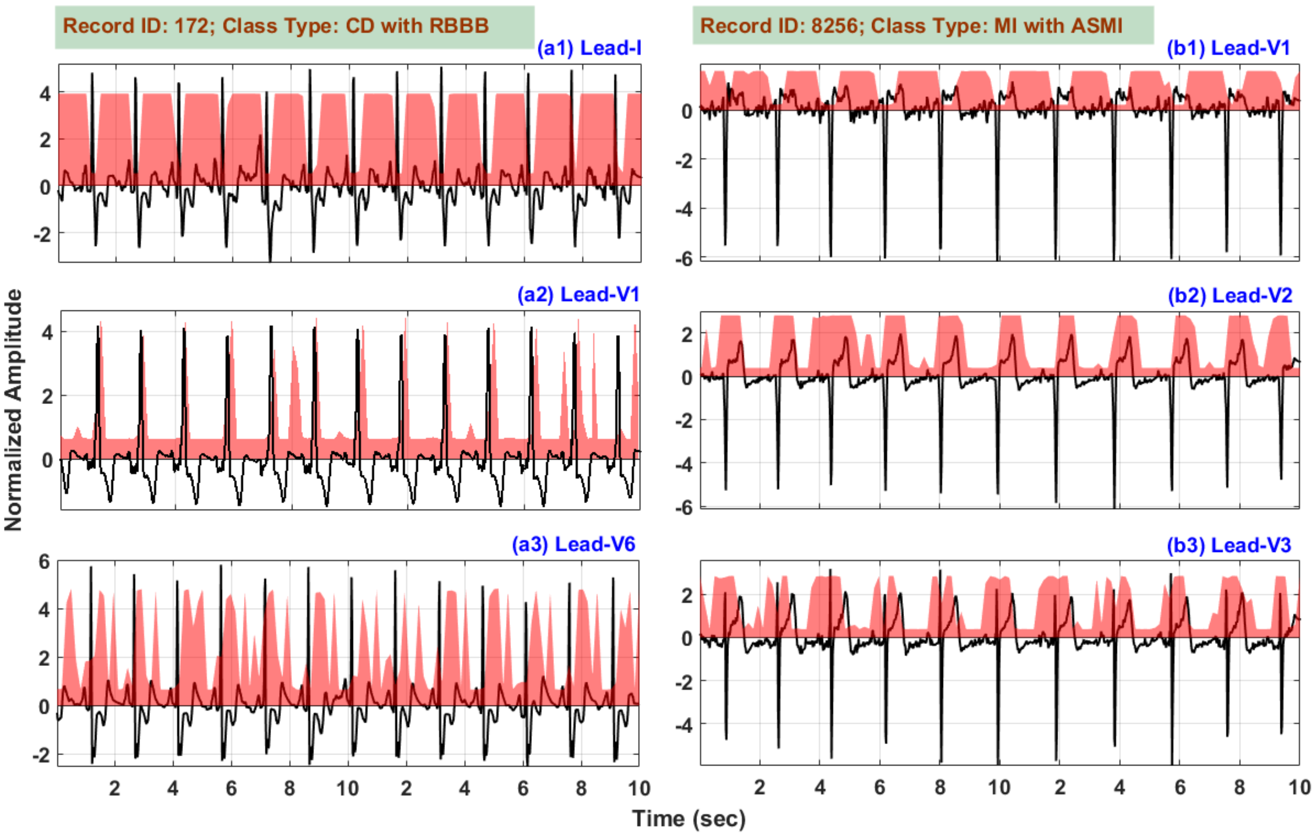


Figure 5.10: Visualization of intra-lead attention weights (red color maps) assigned to the different segments of the ECG records (from the PTBXL-2020 dataset) by the best SD-DTCNN expert in the ensemble. (a1)-(a3) shows s_2 experts' intra-lead attention maps for some critical ECG-leads of the conduction disturbance (CD) patient with complete RBBB abnormality. (b1)-(b3) shows s_1 experts' intra-lead attention maps for some critical ECG-leads of the myocardial infarction (MI) patient with anterior-septal MI (ASMI) abnormality.

region. Figure 5.9 (c)-(d) show the attention maps for two other rhythm ECG records having few abnormal ventricular ectopic beats. It can be observed that the model is highly focused on these abnormal beats over other beats while making a diagnosis decision. Similar attention maps can also be seen for some salient ECG-leads of the CD patient with complete RBBB (Figure 5.10 (a1)-(a3)) and MI patient with anterior-septal MI (Figure 5.10 (b1)-(b3)) from the PTBXL-2020 dataset. It can be observed that for RBBB patient, the abnormal ECG segments including the broad QRS-complexes in lead-V1 (Figure 5.10 (a2)), broad and deep S-waves with ST-T changes in lead-V6 (Figure 5.10 (a3)) and lead-I (Figure 5.10 (a1)) are emphasized by the model. For anterior-septal MI patient, the abnormal ST-elevations and T-waves in the anterior-lead, i.e., V3 and the septal-leads, i.e., V1 and V2, are given more priority by the model. In summary, the proposed model focuses on the clinically relevant ECG segments to make a reliable cardiac abnormality prediction that often aligns with medical knowledge.

Table 5.5: Comparison of number of model parameters

Method	Number of Trainable Parameters PTBXL-2020 Dataset	CinC-2017 Dataset
ATI-CNN [126]	4,998,566	Not applicable
TA-CNN [169]	Not applicable	1,220,000
MS-CNN [118]	12,312,861	12,305,307
DMSFNet [129]	7,656,197	7,652,233
Proposed MS-DTCE	**1,102,225**[a]	**93,339**[b]

[a] SD-DTCNN experts with 12 LS-TCNNs to process 12-lead ECG input. Gating network with 12-lead ECG input.
[b] SD-DTCNN experts with one LS-TCNN to process single-lead ECG input. Gating network with single-lead ECG input.

5.2.7.5 Model Parameters

Table 5.5 compares the number of trainable parameters involved in the design of the proposed model and the best performing baseline architectures [118, 126, 129, 169]. It can be seen that the proposed MS-DTCE consists of fewer training parameters over existing methods and still achieve improved classification results. The fewer training parameters alleviate the risk of over-fitting and improve model generalizability (Table 5.2 and Table 5.3). In order to generate classification decision for the 12-lead ECG input, the best performing single DCNN/DRNN method, i.e., ATI-CNN [126] takes an average test run-time of 4.94 ms and the concatenation-fusion method, i.e., DMSFNet [118] takes 5.81 ms. As expected, deep ensemble methods, i.e., the proposed average ensemble and the MS-DTCE, take higher run-time of 13.20 ms and 15.38 ms, respectively. Although fewer parameters reduce the computational burden for the MS-DTCE, the involvement of multiple experts and processing each lead of the 12-lead ECG separately at each expert increases the average run-time. However, considering the order of input ECG duration in seconds, the MS-DTCE can provide rapid and improved diagnosis decisions.

The improved classification performance of the proposed MS-DTCE method can be attributed to the data-adaptive fusion of diagnosis decisions from the multiple SD-DTCNN experts that exploit key multi-scale pathological ECG variabilities for effective classification. It should be noted that the MS-DTCE method is designed for the multi-class ECG classification, i.e., the method assigns only one target label or disorder to each ECG record. However, in practice, it is common to see multimorbidity patients, i.e., the co-occurrence of two or more acute cardiac disorders such as MI, HYP, and CD, in the same individual [37]. The presence of multiple cardiac disorders simultaneously in an individual could potentially lead to poorer functional status, lower QoL, and increased risk of mortality. Therefore, early detection of multimorbidity patients is crucial to prevent adverse health outcomes and optimize patient-centered care delivery [38].

Considering the difficulty of understanding the complex interaction of multi-scaled pathological ECG features for multimorbidity patients, most researchers have investigated multi-class ECG classification methods. Recently, few researchers [130–133] have presented their preliminary studies on the multi-label ECG classification using a small dataset of 476 multi-label ECG recordings from the CPSC2018 dataset. These methods assign a set of target disease categories to each ECG record, thereby identifying multimorbidity patients. In general, the multi-label classification methods can be divided into two types: problem transformation and algorithm adaptation [171]. Problem transformation methods transform the multi-label classification into a well-known binary classification learning problem. Algorithm adaptation methods adapt the existing learning algorithms to tackle the multi-label classification problem directly. In the direction of multi-label ECG classification, most researchers have employed algorithm adaptation methods [130–133], where they directly train the model with multi-labeled ECG data assuming that the training dataset consists of all the possible label combinations. However, the real-world ECG datasets do not have many label combinations, making the dataset highly imbalanced for different disease label combinations. Improving multi-label ECG classification performance in such scenarios requires proper selection of thresholds for each label, which is burdensome and infeasible. To address the limitations of the existing algorithm adaptation methods, we present a multi-scale deep ensemble framework for multi-label ECG classification using a problem transformation approach in the following section.

5.3 Problem Transformation based Deep Ensemble Approach for Multi-Label ECG Classification

This section presents a generic deep neural network architecture using a problem transformation-based approach for multi-label classification of 12-lead ECG into four cardiac disorders such as CD, HYP, MI, and STTC. Specifically, we transform the complex multi-label classification problem into an ensemble of single-label binary classifiers trained one for each cardiac disorder type. The diagnosis decisions from the binary classifiers are used to detect multiple co-occurring cardiac disorders appearing in the same ECG record. We design each binary classifier using the attention-based temporal convolutional neural network (ATCNN). The ATCNN consists of a stack of temporal convolutional layers followed by intra- and inter-lead attention modules. The temporal layers are designed using dilated convolution filters with different sizes or receptive fields to effectively extract the temporal dynamics of multi-scaled ECG features. As discussed before, the intra- and inter-lead attention modules exploit the 12-lead ECGs within- and across-leads disease-specific diagnostic information, respectively, and enhance the multi-lead feature extraction pipeline. In addition, further analysis of the multi-label diagnosis decisions from the proposed framework can identify low- and high-risk patients assessed in terms of the number of co-occurring cardiac disorders. It is to be noted that we can also utilize the MS-DTCE (multi-class model) to design single-label binary classifiers. However, considering the computational burden of MS-DTCE, employing it for multi-label analysis is infeasible; thus, we design a simple ATCNN model having multi-scaled ECG feature learning ability for the binary classification of single-label disorders.

Figure 5.11 (a) presents an overview of the proposed problem transformation-based deep ensemble approach for multi-label ECG classification. As can be seen, first, the incoming 12-lead ECG input is fed to an ensemble of single-label binary classifiers trained independently, one for each cardiac disorder type. Second, the diagnosis predictions from the binary classifiers are thresholded to identify the multi-label ECG abnormalities. Figure 5.11 (b) illustrates the proposed ATCNN model architecture used to design each disease-specific binary classifiers (disease exists or not). The ATCNN contains multi-lead feature extraction and classification modules. The details of the modules are described below.

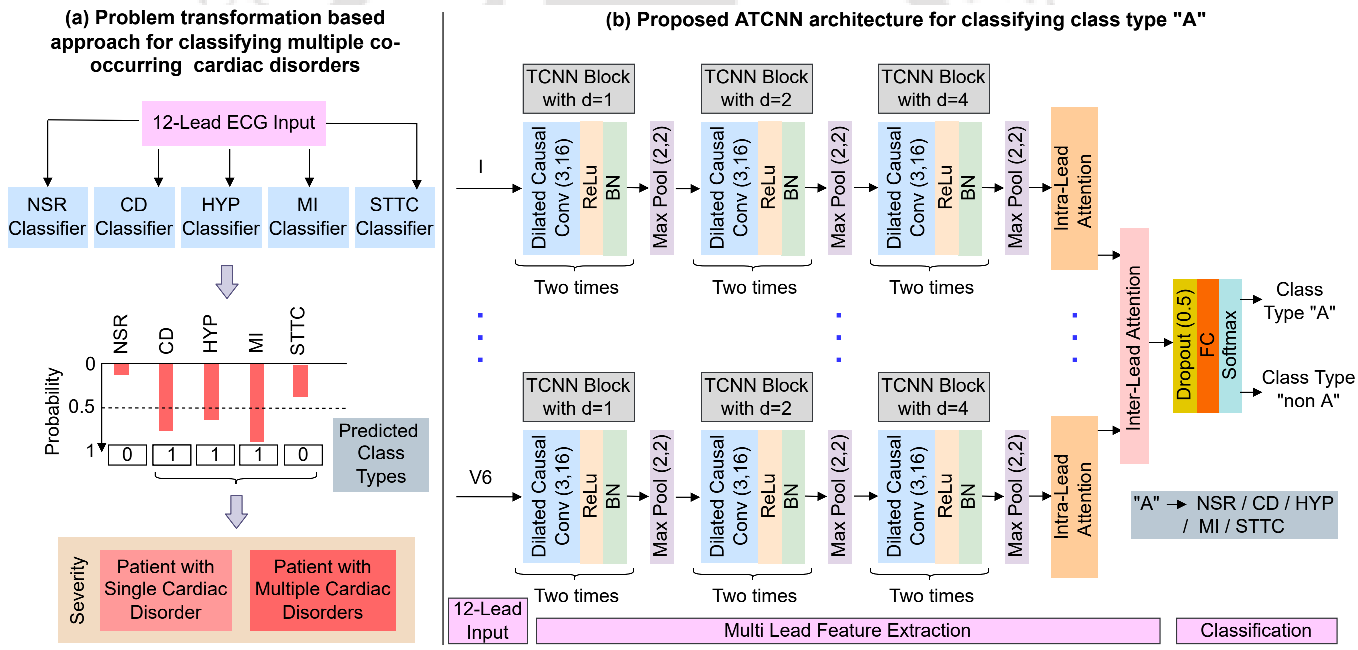


Figure 5.11: (a) Proposed problem transformation-based deep ensemble approach for multi-label ECG classification and severity assessment. (b) Proposed ATCNN model architecture for single-label binary classification of class type "A" using 12-lead ECG signals. The class type "A" can be NSR or CD or HYP or MI or STTC. "Conv(kernel size, number of filters)" and max-pooling window size 2 with stride 2. BN: batch normalization.

5.3.1 Multi-Lead Feature Extraction

This module automatically learns the informative features from the input 12-lead ECG using a stack of temporal CNN (TCNN) layers followed by intra- and inter-lead attention layers. The different leads of the 12-lead ECG exhibit distinct temporal variations in the ECG features that appear at different shapes and morphology. These lead-specific dynamics of multi-scaled ECG features are encoded using a stack of

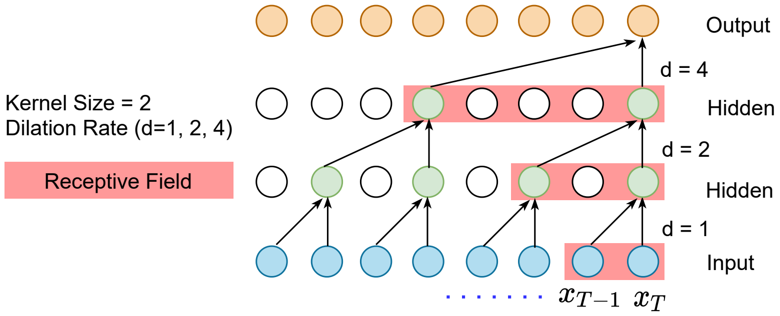


Figure 5.12: Example of dilated and causal convolutions in TCNNs with filter size K = 2 and dilation rate d = [1, 2, 4].

TCNN layers. Each TCNN layer operates at different receptive fields to effectively represent the multi-scaled disease variabilities (Figure 5.11 (b)). The varied receptive field size is attained using dilated and causal convolutional filters that operate at different dilation rates (d). As discussed in the previous section, the effective receptive field size K' of the dilated and causal filter with a dilation rate d and filter size K can be computed as $K' = (K-1)d+1$. Figure 5.12 shows TCNN layers with different dilation rates. As can be seen, the effective receptive field size grows exponentially with the layer depth, which helps encode the multi-scaled ECG variabilities of cardiac disorders. Moreover, as discussed previously, the TCNN layers with dilated and causal convolution filters effectively handle the temporal dependencies of ECG signals for improving diagnosis performance.

Let $\mathbf{X} \in \mathbb{R}^{12\times T}$, $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^l, ..., \mathbf{x}^{12}]^\mathsf{T}$ denote the input 12-lead ECG sequence. Here, T is the transpose operator and $\mathbf{x}^l \in \mathbb{R}^T$ denote the l^{th} ECG lead data with length T. The input 12-lead ECG ($\mathbf{X}$) is fed to the multi-lead feature extraction module to extract the discriminative feature representations. Figure 5.11 (b) illustrates the network configuration of the proposed multi-lead feature extractor consisting of a stack of TCNN layers followed by intra- and inter-lead attention layers. First, each lead of the 12-lead ECG is fed to a stack of TCNN layers that operate at different dilation rates to extract the multi-scaled ECG variabilities. The output feature maps of the TCNN layers for the l^{th} ECG lead ($\mathbf{x}^l$) can be given as:

$$\mathbf{F}_l = f_{d=4}^{TCNN}\left(f_{d=2}^{TCNN}[f_{d=1}^{TCNN}(\mathbf{x}^l)]\right) \tag{5.13}$$

where f_d^{TCNN} represents a TCNN block consisting of two TCNN layers with dilation rate d and a stack of such blocks maps the l^{th} ECG lead ($\mathbf{x}^l$) to a output feature maps $\mathbf{F}_l \in \mathbb{R}^{N_f \times T_f}$, with $l = 1, 2, ..., 12$. Here, N_f and T_f represents the number of feature maps and their length, respectively. Next, the lead-specific feature maps $\mathbf{F}_l$ are fed to the intra- and inter-lead attention layers. These layers emphasize/alleviate the within- and across-lead clinically relevant/redundant ECG information and generate a high-level across-lead attentive representation (ALAR) vector $\mathbf{a} \in \mathbb{R}^{N_f}$ used for the classification as:

$$\mathbf{a} = \mathcal{F}_{Inter}\left[\mathcal{F}_{Intra}(\mathbf{F}_1), \mathcal{F}_{Intra}(\mathbf{F}_2), ..., \mathcal{F}_{Intra}(\mathbf{F}_{12})\right] \tag{5.14}$$

where $\mathcal{F}_{Intra}$ and $\mathcal{F}_{Inter}$ denote the intra- and inter-lead attention layers functions, respectively. The mathematical formulations of these layers are similar to that discussed in chapter 2.

5.3.2 Single-Label Binary Classification

The high-level ALAR vector ($\mathbf{a}$) is fed to the fully-connected output layer with Softmax activation, defined as Softmax(x_i) = $\frac{1}{Z}$ exp(x_i) with $Z = \sum_i$ exp(x_i), to compute the probability distribution of outputs for the input 12-lead ECG $\mathbf{X}$ and is given as:

$$\begin{bmatrix} p(A|\mathbf{X}) \\ 1 - p(A|\mathbf{X}) \end{bmatrix} = Softmax(\mathbf{W}_o\mathbf{v} + \mathbf{b}_o) \tag{5.15}$$

where $p(A|\mathbf{X})$ and $(1-p(A|\mathbf{X}))$ represent the probability of $\mathbf{X}$ belongs to class type A and $nonA$, respectively, and $A \in \{NSR, CD, HYP, MI, STTC\}$ (Figure 5.11 (b)). $\mathbf{W}_o$ and $\mathbf{b}_o$ are weights and biases.

5.3.3 Training Model Parameters

The network parameters of the single-label binary classifier are trained or optimized using the binary cross entropy loss function ($\mathcal{L}$) computed as:

$$\mathcal{L} = ylog(p(A|\mathbf{X})) + (1 - y)log(1 - p(A|\mathbf{X})) \tag{5.16}$$

where y represents the true label of the input $\mathbf{X}$, $y \in \{0, 1\}$ with 1 for class type A and 0 for type $nonA$.

5.3.4 Multi-Label Classification Strategy

After training a set of binary classifiers one for each class type, the multi-label ECG classification is performed as follows: first, the input $\mathbf{X}$ is fed to the five single-label binary classifies and obtained the probability of output class prediction as $[p(NSR|\mathbf{X}); p(CD|\mathbf{X}); p(HYP|\mathbf{X}); p(MI|\mathbf{X}); p(STTC|\mathbf{X})]$. Finally, multiple co-occurring cardiac disorders (multi-label) in the same ECG record are determined with 0.5 thresholding on the probability of predicting a specific class type (Figure 5.11 (a)). In addition, analysis of the multi-label diagnosis decisions from the proposed framework can identify low- and high-risk patients assessed in terms of the number of co-occurring cardiac disorders, i.e., patients with a single cardiac disorder and patients with multiple co-occurring disorders, respectively (see Figure 5.11 (a)).

Table 5.6: Distribution of ECG recordings for the five diagnostic labels (NSR, CD, HYP, MI and STTC) and their combinations across the development and the test sets of the PTBXL-2020 database.

Label Type	Development	Test	Concurrent Labels	Development	Test
NSR	8093	902	One-label	14465	1638
CD	4054	453	Two-label	3335	349
HYP	2494	271	Three-label	849	100
MI	4933	553	Four-label	149	16
STTC	4704	519	Five-label	0	0
Total	24278†	2698†	**Total**	18798	2103

†: More number of records than original due to the multi-label annotations of some ECG records.

5.4 Results and Discussion for the Multi-Label ECG Classification

This section presents the details of the clinical ECG datasets, the signal preprocessing, and the evaluation metrics employed. This section also presents the details of the network optimization and the experimental results of the proposed problem transformation-based deep ensemble approach for multi-label ECG classification, followed by the performance comparison with the existing methods.

5.4.1 Clinical ECG Database

In this study, we have used the PTBXL-2020 dataset for model development and evaluation. The PTBXL-2020 is the largest publicly available multi-label 12-lead ECG dataset comprising 21,837 records from 18,885 patients (females: 9,064; males: 9,821) sampled at 100 Hz for 10 s duration. The database provides multi-label annotations for each 12-lead ECG record in terms of NSR or normal, CD, HYP, MI, and STTC. Also, the database provides the development and test partitions of the ECG records. The ECG records which do not have proper diagnosis annotations are excluded from the study. Table 5.6 shows the distribution of records across the five diagnostic labels and their multi-label combinations for the development and test sets. As can be seen, around 23% of records (4,798 out of 20,901) have more than one cardiac disorder at the same time; this emphasizes the importance of multi-label ECG classification.

5.4.2 Dataset for Training Binary Classifiers

This study employs a problem transformation approach for multi-label ECG classification; thus, the multi-label dataset is divided into multiple sub-datasets representing each cardiac state. For example, the sub-dataset of MI consists of MI and non-MI class ECG records. Here non-MI class consists of ECG records from all other cardiac states except MI. It is observed that the sub-datasets of each cardiac state (except NSR class) show an imbalanced ECG data distribution, i.e., the number of positive class samples is less than the negative class; thus, a random undersampling algorithm is employed to mitigate class imbalance. For example, to train a binary classifier for the MI (MI vs. non-MI), the non-MI ECG recordings are selected randomly from the development dataset (see Table 5.6) until twice as the MI recordings. The selected sub-dataset is randomly divided into training and validation sets in the ratio of 8:2. It is to be noted that the random undersampling is applied only on the development dataset, and the test dataset is exclusively used for evaluating the performance of the proposed ATCNN model. The ECG signals are preprocessed to remove baseline noise, followed by z-score normalization.

5.4.3 Model Configuration, Training Settings and Performance Measures

The proposed ATCNN configuration is described in Figure 5.11 (b). The hyperparameters, such as the number of TCNN blocks and the convolutional filter size, are set experimentally. The model is trained with an Adam optimizer with a learning rate of 0.001 on mini-batches of size 32 for 100 epochs. It is to be noted that the same ATCNN architecture is used for each cardiac condition (NSR/CD/HYP/MI/STTC), and their respective model parameters will be optimized during the training process. The classification performance of each cardiac condition is evaluated by the standard binary classification measures such as recall (Re), precision (Pr), specificity (Sp), accuracy (Acc), and F1-score. In addition, an exact match measure is also employed to evaluate the direct match between the predicted and true multi-label vectors, ignoring the subset of correctly predicted labels.

5.4.4 Multi-Label ECG Classification Results

Table 5.7 shows the class-wise and average classification performance of the proposed ATCNN model on the test dataset. As can be seen, the model achieved an average Re of 83.88%, Pr of 70.80%, Sp of 89.01%, Acc of 88.05%, F1-score of 76.51%, respectively. Specifically, more than 81% of Re across all cardiac conditions shows that the proposed model has good sensitivity for detecting multiple cardiac disorders. Although the pathological ECG features that characterize CD, MI, and STTC may be subtle and similar at certain ECG-leads, the model still achieved a promising F1-score of 77.18%, 76.41% and 76.34%, respectively. However, the under-representation of HYP samples in the training dataset slightly degrades its F1-score on the test dataset. In addition, most of the ECG records concerning the presence or absence (TP and TN) of various cardiac disorders are correctly classified with fewer misclassified (FP and FN) records. The proposed multi-label ECG classification method achieves a multi-label exact match measure of 66.38% on the test data (Table 5.7). The experimental results of the proposed model demonstrate good detection performance for each cardiac condition. An ensemble of single-label binary classifiers with a

Table 5.7: Proposed ATCNN model performance on the PTBXL-2020 test set. TP: true positives, FN: false negatives, FP: false positives, and TN: true negatives.

Class	TP	FN	FP	TN	Re(%)	Pr(%)	Sp(%)	Acc(%)	F1-score(%)
NSR	811	93	172	1027	89.71	82.50	85.65	87.40	86.01
CD	367	86	131	1519	81.01	73.69	92.06	89.68	77.18
HYP	230	41	190	1642	84.87	54.76	89.62	89.01	66.57
MI	452	101	178	1372	81.73	71.74	88.56	86.73	76.41
STTC	426	93	171	1413	82.08	71.35	89.20	87.44	76.34
				Average	**83.88**	**70.80**	**89.01**	**88.05**	**76.51**
Multi-label exact match(%)					**66.38**				

fixed threshold of 0.5 on their predictions yields promising results for multi-label ECG classification. In practice, grouping patients with a single cardiac disorder and multiple co-occurring disorders is critical in timely initiating life-saving treatment plans for high-risk multimorbidity patients. Thus, the decisions from the proposed framework are further analyzed to identify single and multimorbidity patients (Figure 5.11 (a)). The method correctly identified 408 single cardiac disorder patients out of 736 and 177 multimorbidity patients out of 465 from the test dataset. We report these results by matching true and predicted label vectors, ignoring the subset of correctly predicted labels.

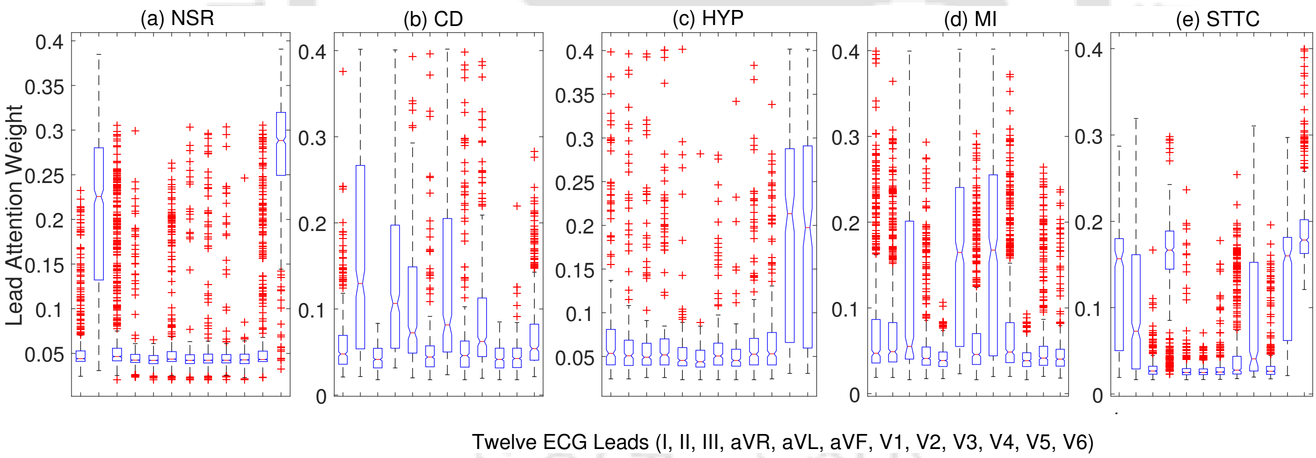


Figure 5.13: Boxplot of the inter-lead attention weights learned by the ATCNN model for 12-lead ECG illustrate the importance of each ECG-lead for diagnosing (a) NSR, (b) CD, (c) HYP, (d) MI and (e) STTC on the validation dataset.

5.4.5 Model Interpretability

Figure 5.13 (a)-(e) show the distribution of inter-lead attention weights for the 12-lead ECG records of NSR, CD, HYP, MI, and STTC, respectively. As can be seen, the proposed model provides different weights of importance for each ECG-lead based on their diagnostic relevance or redundancy in identifying the specific cardiac condition. Specifically, for the CD class, ECG-leads II, aVR, V1, aVL, and V3 contribute

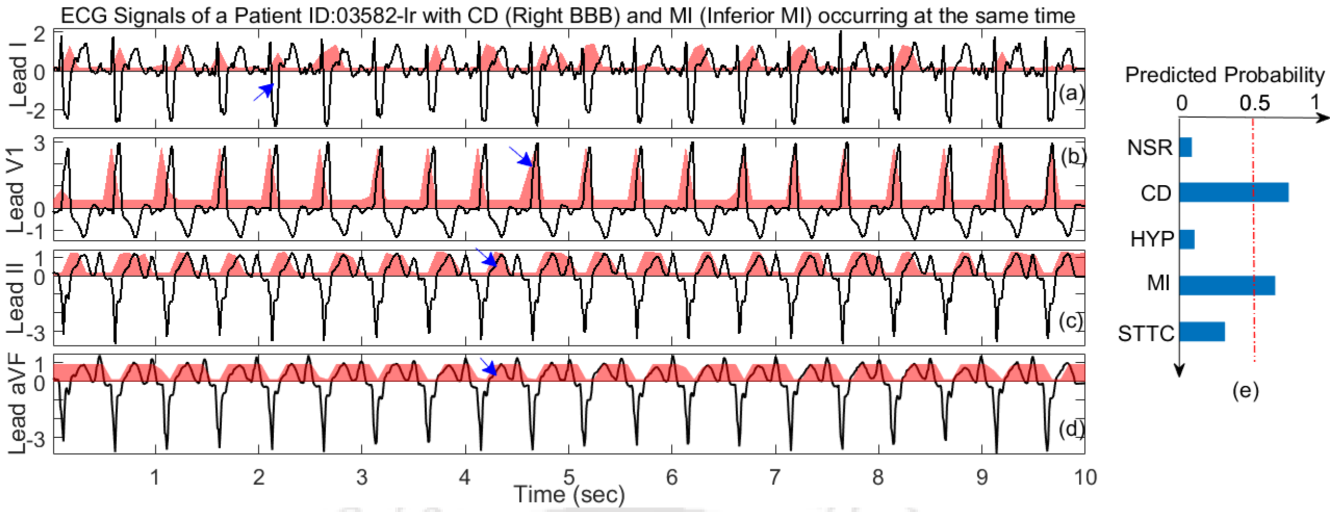


Figure 5.14: Visualization of intra-lead attention weights (red area height) assigned for different segments in salient ECG-leads of a patient ID:03582-lr with RBBB and inferior MI occurring at the same time (example from the test dataset). (a)-(b) Attention maps generated by the CD class binary classifier for ECG-leads I and V1. (c)-(d) Attention maps generated by the MI classifier for ECG-leads II and aVF. As highlighted by the blue arrow, the red area height is more for the pathological ECG features to some extent. (e) The proposed approach correctly identifies the co-occurring cardiac disorders using a simple threshold of 0.5 on each class prediction probability.

more to the classification. Similarly, ECG-leads V5, V6, I, and V4 for HYP; V2, aVF, III, II, and V3 for MI; and V6, aVR, V5, I, II, and V3 for STTC class. The reason for giving more weight to these ECG-leads is that clinicians often examine the abnormal QRS-morphology in leads V1 and aVL to diagnose CDs such as right BBB. The large R-wave in lateral leads V5, V6, and I helps detect HYP disease. Inferior leads II, III, and aVF and anterior leads V3 and V2 are used in the clinic to detect the ST-elevation inferior and anterior MI, respectively; the ST-depression and T-changes in lateral leads V5, V6, and I are used in clinics to identify the lateral wall ischemia. Although ECG-lead aVR is often ignored in clinics, it helps diagnose CD, and STTC patients [172]. Figure 5.14 (a)-(d) shows the visualization of intra-lead attention weight maps for a key ECG-leads of a patient with complete RBBB and inferior MI occurring at the same time. The proposed approach correctly identified these co-occurring disorders using simple thresholding on the prediction probability of single-label binary classifiers (Figure 5.14 (e)). The attention weight maps for the optimal ECG-leads of RBBB (Figure 5.14 (a)-(b)) and inferior MI (Figure 5.14 (c)-(d)) disorders are extracted from their corresponding binary classifier. As can be seen, the intra-lead attention layer provides different weights for different ECG features based on their relevance in the diagnosis process. In practice, clinicians examine (i) the deep S-waves in the lead I and broad QRS-complexes in the lead V1 for diagnosing RBBB disorder and (ii) the distinct ST-segment elevations in the inferior leads II and aVF for diagnosing inferior MI. Interestingly, the attention maps learned by the model indeed provide more weight to these pathological ECG features during the diagnosis process. In summary, the intra- and inter-lead attention weights generated by the proposed model match with the clinical evidence to some extent, which aided in improved performance and promising model interpretability results.

Table 5.8: Comparison of the proposed method with existing multi-label ECG classification methods.

Reference	Method	Database (#Records)	#Multi-Label Records	Performance
[130]	117 handcrafted feature + Multi-objective classifier	CPSC2018 (6,877)	476	F1-score = 60.8%
[131]	Sequence generation-based DCNN model	CPSC2018 (6,877)	476	F1-score = 87.2%
[132]	DCNN model	CPSC2018 (6,877)	476	F1-score = 96.7%
[133][†]	CNN + BiLSTM + Attention layers	CPSC2018 (6,877)	476	F1-score = 60.1%
Proposed	TCNNs with different receptive fields + Attention layers	PTBXL-2020 (**21,836**)	**4,798**	F1-score = **76.5**%

[†]: This method uses single-lead ECG input, and all other methods use 12-lead ECG input. #: Number of.

5.4.6 Comparison with Existing Methods

Table 5.8 compares the performance of the proposed and existing multi-label ECG classification methods. Existing methods [130–133] mostly used algorithm adaptation approaches for multi-label ECG analysis. However, these methods require complex models with rigorous threshold adjustment to achieve good multi-label ECG classification performance. On the other hand, the proposed method is based on the problem transformation approach, where the complex problem is transformed into a set of simple single-label binary classifiers. An ensemble of predictions from these classifiers provided promising performance for multi-label ECG classification with an average F1-score of 76.5% (Table 5.8). Existing methods [130–133] employed a small CPSC2018 dataset containing only 476 multi-label 12-lead ECG records for the evaluation. Thus, the higher F1-score of the methods [131, 132] may be attributed to the correct detection of records with single-label disorders. On the other hand, the proposed method is validated on a large PTBXL-2020 dataset consisting of 4,798 multi-label ECG records. Moreover, the dataset contains ECG records of patients with acute disorders such as MI, HYP, and CD; thus, their co-occurrence detection helps clinicians identify high-risk patients and initiate timely treatment. The proposed DL-based model also shows promising model interpretability results; thus, it may help clinicians and researchers to understand the salient ECG-leads and pathological ECG features that led to the diagnosis. Finally, compared to the existing methods, incorporating TCNN layers with different dilation rates followed by attention layers in the proposed model effectively learned the multi-scaled pathological ECG features dynamics, improving classification performance. In order to generate the multi-label classification decision for the 12-lead ECG input, the proposed framework takes an acceptable average test run-time of 9.59 ms.

5.5 Summary

This chapter explored the use of multi-scale temporal CNN ensemble frameworks for the automated analysis of multiple acute cardiac disorders from 12-lead ECG. Specifically, the MS-DTCE model is proposed for effective multi-class ECG classification. The main contributions include; (i) the design of multiple SD-DTCNN expert classifiers that exploit key multi-scaled pathological ECG variabilities for generating local predictions and the design of a gating network for combining the experts' local predictions in a data-adaptive manner to generate the final diagnosis decision; (ii) the formulation of the new error function with a correlation penalty for optimizing various modules of proposed architecture simultaneously in an end-to-end manner. Experimental results demonstrated that our approach effectively classifies cardiac disorders over baseline methods by combining diagnosis decisions from multiple experts. The proposed MS-DTCE has the advantages of high diagnostic accuracy for a broad range of acute cardiac disorders, elegant generalization ability for effectively handling ECG signals with different leads (12-leads/single-lead), and encouraging model transparency results; thus, it has the potential to be employed for reliable cardiac ailments screening applications.

In this chapter, we also presented the use of a problem transformation-based deep ensemble approach for the multi-label ECG classification. Specifically, a set of ATCNN-based single-label binary classifiers are trained for each cardiac disorder, and these classifiers can extract the multi-scaled ECG features using TCNN layers and attention modules for improved performance. The predictions from the binary classifiers with simple thresholding generated the final multi-label decisions. The model design is more efficient than the algorithm adaptation methods for multi-label ECG analysis and can be extended to any additional disorders with relatively minimal effort. The preliminary findings suggest that transforming the multi-label problem into an ensemble of single-label classifiers demonstrated promising detection for co-occurring cardiac disorders appearing in the same ECG record. The experimental results also demonstrated the advantage of this approach for identifying single and multimorbidity patients; thus, the method has the potential to screen patients at high-risk and help clinicians initiate timely treatment.

6

Conclusions

Contents

6.1 Summary of the Work

According to the American Heart Association/American College of Cardiology (AHA/ACC), the global prevalence of heart diseases is rising at an alarming rate, from 271 million in 1990 to 523 million in 2019, and it is estimated to increase exponentially by 2030 [3]. The recent projections show that the high prevalence of heart diseases will result in a massive shortage of physicians per patient, an increase in overall costs for patient care, and mortality rates worldwide [16]. It would be troublesome for any cardiac clinics to provide screening services at a large scale with a limited number of experienced medical practitioners. Unfortunately, recent evidence suggests that the risk and 1-year burden of heart disease in survivors of acute COVID-19 are substantial [173], which further adds to the increase in prevalent cases. Consequently, there is an urgent unmet need for automated ECG interpretation methods as they can accomplish fast and reliable screening for cardiac disorders in home-care or hospital-care applications, improve disease management, and reduce the overall cost of care. In addition, most cardiac disorders progressively worsen over time. Thus, developing intelligent automated methods suitable for portable remote health monitoring ECG devices can enable early diagnosis and severity assessment of cardiac disorders, thereby aiding clinicians in initiating appropriate treatment strategies that could help avert further disease progression and improve patient outcomes. In this direction, the current thesis presented various automated ECG interpretation methods for the diagnosis and severity assessment of cardiac disorders using 12-lead and single-lead ECG signals. The methods exploited the pathological ECG manifestations of cardiac disorders and effectively handled the disease variabilities to develop reliable and robust DL-based automated approaches. The summary of each chapter of the thesis is given as follows.

In Chapter 1, the physiological origin of the ECG signal and the popularly used ECG recording systems, including the standard 12-lead ECG and the single-lead ECG, have been introduced. This chapter also discusses various life-threatening cardiac disorders and their associated pathological ECG manifestations, followed by the challenges encountered during the clinical diagnosis. Subsequently, the literature review of the existing automated methods for diagnosing cardiac disorders from single- or 12-lead ECG signals is presented. Although many automated ECG interpretation methods are proposed for diagnosing cardiac disorders, few research gaps existed that were not addressed effectively. This chapter briefly discusses these research gaps and the motivation for the current research in the last section. The following paragraphs summarize the major findings related to each motivation and list the most important conclusions.

Although MI is a progressive cardiac disorder, existing methods [69, 71–78] consider it as one broad category for the diagnosis and do not address its severity staging. The subtle variabilities in the pathological

ECG manifestations and dynamic changes in their temporal dependencies during the disease progression pose a challenge for reliable classification of MI severity stages (early MI, acute MI, and chronic MI). In addition, the implicit diagnostic redundancy of the 12-lead ECG often obscures the clinically relevant information, which further adds to the difficulty of classifying severity stages. To address the above challenges, in Chapter 2, we presented an MLDA-RNN method for the reliable automated classification of MI severity stages from 12-lead ECG. The model systematically processes the 12-lead ECG using lead-specific RNNs followed by intra- and inter-lead attention modules. Specifically, the lead-specific RNNs effectively characterize the temporal dynamics of subtle pathological manifestations associated with the MI severity. The subsequent intra- and inter-lead attention modules automatically emphasize/alleviate the diagnostically relevant/redundant within- and across-lead information during the feature fusion to obtain the discriminative feature representations for reliable classification. The experimental results verified that the combination of RNNs and intra- and inter-lead attention modules effectively combines the relevant within- and across-lead diagnostic information of the 12-lead ECG, which aided in an impressive OA of 97.79% for severity stages classification. Along with the improved performance, the notable advantage of the MLDA-RNN method was the diagnostic transparency provided by the attention weights. Visualizing intra- and inter-lead attention weights verified the correlation of weights with the diagnostic relevance of focused ECG features and ECG leads, which often aligns with medical knowledge and correlates with the cardiologist's way of 12-lead ECG examination. In this chapter, we also presented the effectiveness of fusing the patient's clinical features with the 12-lead ECG to improve MI and non-MI diagnosis in the presence of mimicking MI patients. The preliminary findings suggested that incorporating the patient's clinical history and the 12-lead ECG into MI evaluation improved OS by 7.5% over 12-lead ECG alone.

Chapter 3 addressed the task of automated CHF diagnosis from the single-lead ECG heartbeats. The diverse etiology of CHF results in wide variability in the pathological ECG features and dynamic changes in their temporal dependencies. The existing methods employ DCNN-based feature representation models [49, 88, 89], which may not adequately capture complex pathological ECG features and their temporal dependencies associated with the CHF syndrome. Moreover, these deeper networks are hard to optimize because of the vanishing gradient problem, and also, these methods lack "diagnostic transparency," which limits their applicability in real-world clinical applications. To tackle these issues systematically, we presented the A-DRRNet method that attentively captures the complex pathological ECG features and their temporal dependencies (P-QRS-T) associated with the CHF syndrome to make the diagnosis decision and enhance the model transparency. Specifically, the proposed method employed multi-layered RNNs consisting of recurrent hidden states to encode the temporal dynamics of complex pathological ECG features effectively. In addition, residual connections have been introduced between the RNN layers to ease the training of the proposed deeper network, which we have shown with mathematical derivations. Finally, we incorporated an attention module to further improve feature learning by focusing on clinically relevant ECG information for efficient CHF detection. The proposed method improved the diagnostic accuracy of CHF detection by nearly 4.6$\%$ compared to the existing methods, achieving an Acc of 98.57$\%$. Notably, the advantage of the proposed method was the diagnostic transparency achieved by visualizing the attention weight heat maps. The visual analysis of the weights and the corresponding ECG features verified that the A-DRRNet model learned to focus on ECG features that characterize CHF for the diagnosis, which imparted adequate interpretability to the model and made it more reliable for clinical applications.

AF is the most common heart rhythm disorder that can progressively worsen over time and increases the risk for ischemic stroke five-fold. Because AF can be intermittent and asymptomatic, it remains undiagnosed with the conventional snapshot 30 s ECG evaluation performed during routine hospital visits. As a result, ischemic stroke was often the first clinical presentation of AF in many patients. It is thus a disorder that would benefit from continuous monitoring with Holter ECG devices. Analyzing AF burden from these long-term ECG recordings can aid in the early diagnosis and severity assessment of AF condition and help assess the patient's response to stroke prevention therapies. Existing approaches [90, 113–118] focus on detecting the presence or absence of AF and do not quantify the disease severity via AF burden. The precise AF burden estimation from the long-term recordings is often challenging due to frequent ectopic beats, i.e., AF-masquerading rhythms and different noise levels. To address the above challenges, in Chapter 4, we presented an MT-DCNN method for accurate and robust AF burden estimation from the single-lead long-term ambulatory ECG recordings. The method included AF detection as a primary task and the ECG sequence reconstruction as an auxiliary task using DCNNs. Specifically, we incorporated a CDAE-based auxiliary task, which was shown to regularize the MT-DCNN model to learn robust feature representations for efficient AF detection and aided in accurate AF burden estimation from long-term recordings. The experimental results demonstrated that the MT-DCNN model outperformed benchmarked feature engineering and DL approaches with a lesser mean E_{AF} error of 2.8$\%$ on the test set. In addition, the MT-DCNN also demonstrated robust AF burden estimation results on independent test sets with frequent ectopic beats and different noise levels over the baseline methods. Finally, apart from helping identify patients with low- and high-risk for stroke, the analysis of AF burden from the long-term recordings showed an improvement of nearly 20.7$\%$ in AF diagnosis Acc over traditional snapshot ECG evaluation.

In real-world clinical applications, patients can present with various acute cardiac disorders, including MI, BBBs, and HYP. Thus, the automated diagnosis systems should be able to analyze 12-lead ECG signals of multiple cardiac disorders simultaneously and correctly identify the underlying cardiac ailment/ailments for their appropriate management. However, it is challenging to classify multiple cardiac disorders reliably, considering their subtle, similar, and multi-scaled pathological ECG morphologies. Existing single expert/classifier-based approaches [124–129] may not adequately represent the highly variable and multi-scaled ECG manifestations associated with multiple cardiac disorders. In practice, clinicians often seek a second or more opinion for patients with confusing ECG biomarkers before making the final diagnosis. Thus, the weighting of suggestions from several experts can generate a reliable final decision. Inspired by this mixture of expert-based clinical diagnosis, in Chapter 5, we presented multi-scale deep temporal CNN ensemble approaches for effective multi-class and multi-label ECG classification. Specifically, the MS-DTCE method was developed to extract and combine diverse multi-scaled ECG features for effective multi-class ECG classification. The method included several temporal CNN-based expert classifiers (SD-DTCNN) with different receptive fields to extract the temporal dynamics of scale-specific pathological ECG characteristics and generate local predictions. A temporal CNN-based gating network (DTCNN) was designed to aggregate the local predictions of experts based on their competencies. An objective function was also formulated to facilitate the interaction and optimal diversity between the experts, enabling them to learn specialized discriminative features for improved classification. The experimental results on the PTBXL-2020 and the CinC-2017 datasets demonstrated that the MT-DTCE effectively classifies cardiac disorders over baseline methods by combining diagnosis decisions from multiple experts and achieved an impressive F1-score of 84.5% and 88.3%, respectively. The MS-DTCE has the advantages of high diagnostic accuracy for a broad range of acute cardiac disorders, elegant generalization ability for effectively handling ECG signals with different leads (12-leads/single-lead), and encouraging model transparency results. In this chapter, we also presented a problem transformation-based deep ensemble approach for the multi-label ECG classification. Particularly, a set of ATCNN-based single-label binary classifiers were trained for each cardiac disorder using temporal CNN layers with different dilation rates and attention modules for learning multi-scaled ECG variabilities. The predictions from the binary classifiers with simple thresholding generated the final multi-label decisions. The preliminary findings suggest that transforming the multi-label problem into an ensemble of single-label classifiers demonstrated promising multi-label ECG classification performance with an average F1-score of 76.5%. The experimental results also demonstrated the advantage of this approach in identifying single- and multi-morbidity patients.

Overall, the findings of this thesis conformed that the DL methods applied for ECG analysis could improve and extend current diagnostic models, bring new research opportunities and contribute to developing automated decision support systems for home care and hospital care applications.

6.2 Future Directions

In this thesis, we have presented several automated methods for improving the diagnosis of cardiac disorders from the ECG signals. The following would be the directions for future works that might further improve the clinical decision-making process for cardiac ailments.

- Although the standard 12-lead ECG helps diagnose various cardiac abnormalities, it is unreliable for diagnosing posterior cardiac ailments, such as ST-elevation posterior myocardial infarction (PMI). This is because the 12-lead ECG measurement does not have posterior monitoring electrodes. Clinically, ST-depressions in the anterior ECG leads (V1-V4) of 12-lead ECG are examined for PMI diagnosis. However, the ST-depressions in these ECG leads are not specific to PMI as the anterior ischemia also shows similar ST-depressions, which results in poor diagnosis performance. The 3-lead vectorcardiogram (VCG) is an alternative recording system consisting of an electrode monitoring the posterior heart wall; thus, it is more reliable for diagnosing posterior cardiac abnormalities. The 3-lead VCG is acquired using seven electrodes placed in specific locations on the human torso that captures the cardiac electrical activity in three orthogonal directions, namely V_x, V_y, and V_z. As preliminary work [174], we presented an ML-based method that exploits the 3-lead VCG signals using multiscale eigen features followed by a cost-sensitive SVM classifier for identifying PMI patients. Experimental results on a small dataset suggest that 3-lead VCG can significantly improve the PMI detection performance over the 12-lead ECG. However, validation on large datasets is required to verify its clinical utility, which could be a promising direction to be explored in future works.
- In Chapter 2, the MLDA-RNN is proposed for classifying MI severity stages (EMI, AMI, and CMI) using 12-lead ECG. The early stages of MI can be reversible, and if left untreated, it causes severe irreversible damage to the myocardial tissues. Recent studies show the interaction of various risk factors (age, gender, high blood lipids, smoking, diabetes, hypertension, obesity, stress, alcohol intake, and exercise) in accelerating MI progression [152, 175]. Several risk computation models, including TIMI (Thrombolysis in Myocardial Infarction) and GRACE (Global Registry of Acute Coronary Events), employ some of the above risk factors to assess the MI severity [175]. Combining the ECG features from the MLDA-RNN with the patient's risk factors can play a crucial role in designing robust severity assessment tools for MI and help identify patients benefit from the therapy. Although preliminary findings from Chapter 2 demonstrate the effectiveness of combining features for improving MI detection in the presence of mimicking MI diseases, these results require further validation on

large patient cohorts to be used for the risk assessment, which needs to be addressed in the future. Finally, the inter-lead attention weights learned by the MLDA-RNN method can provide the number of salient ECG leads showing pathological manifestations, which can be further analyzed as a risk factor to assess the location, spread, and severity of the infarcted heart region.

- CHF is a chronic progressive condition with different severity levels, such as NYHA types I to IV based on symptoms, and with different subtypes, such as HF with preserved ejection fraction, HF with reduced ejection fraction, and HF with mid-range ejection fraction based on left ventricular ejection fraction (LVEF) measurements. In Chapter 3, we have addressed only CHF detection as a broad category from ECG beats using an A-DRRNet model. Considering CHF is a syndrome, the surface ECG may not provide sufficient evidence for its staging; thus, the proposed approach is only suitable for the preliminary screening of CHF patients. However, the combined analysis of B-type plasma Natriuretic Peptides (BNP) and advanced echocardiogram (ECHO) for the suspected CHF patients (detected using ECG) can provide crucial details about the disease severity and helps choose an appropriate treatment [176], which could be a promising direction to be explored in the future.

- The MT-DCNN model presented in Chapter 4 has shown promising results for AF burden estimation from the long-term Holter ECG recordings over the baseline approaches. However, in the presence of high proportion ectopic beats (LTNSRDB test set), the model predicts more false AF rhythms, thereby increasing E_{AF} error (see Table 4.5). Including more representative data for the ectopic beat rhythms during training would increase the generalization performance and potentially reduce the E_{AF} error. Although the selection of 30 s ECG excerpts provided better results for the current datasets, it may cause slight inaccuracies in AF burden estimation for some cases of paroxysmal AF, in which AF only lasts for a couple of seconds. Further investigations on the optimal analysis window could be one more promising direction to explore. The use of a 4% AF burden threshold for screening low- and high-risk AF patients is motivated by the recent research guidelines [135, 156]; further insights can be explored to determine the best threshold for stroke risk-stratification. In addition, we have employed only AF burden and did not consider the patient's risk profile represented by the CHA2Ds2-VASc score [177] for the stroke risk-stratification. Recent studies [136, 155] have suggested that combining AF burden measures with the patient's risk profile helps design more robust personalized stroke risk-stratification approaches for AF patients, which demands further exploration.

- In the direction of multiple cardiac disorders classification, Chapter 5 investigates the MS-DTCE and the ATCNN ensemble frameworks for multi-class and multi-label ECG classification, respectively.

Specifically, considering the limited multi-label ECG records, we employed a problem transformation-based ATCNN approach for classification. This multi-label method does not consider the label correlation information during analysis. Including more representative label combinations data can help exploit the label correlation information, further improving the multi-label classification performance. Although preliminary findings suggest the effectiveness of the proposed multi-label method, validation on large-scale datasets is required to verify its clinical utility. Besides limited ECG records, using 12-lead ECG alone to detect multiple cardiac ailments has systematic limitations in providing differential diagnoses for mimicking diseases, limiting the performance of multi-class and multi-label ECG classification (see Table 5.2 and Table 5.7). Hence, multimodal analysis of advanced cardiac imaging tools such as MPI, angiography, ECHO, and MRI for the suspected individuals (detected using ECG) could improve the performance further by accurately identifying the origin of the pathology for various cardiac ailments and aid in definitive treatment, which is a promising direction to be explored in the future.

- It is well known that DL models require large amounts of annotated datasets to generate reliable diagnosis results. However, considering the asymptomatic nature of AF and CHF (NYHA I and II), the large-scale datasets of these disorders are less abundantly available. In addition, the lack of large datasets for the early stages of MI and other cardiac disorders (BBBs and HYP) affects the classification performance of the automated methods. For example, insufficient data for these disorders result in a drop in their detection performance, as seen in Table 2.5, Table 5.2, Table 5.3, and Table 5.7. Although cost-sensitive learning and data-augmentation techniques have been adapted, there is still room for improvement. Recently, data-efficient DL approaches such as semi-supervised generative adversarial networks (GANs) have provided better generalizable results in limited training data conditions [178,179]. A detailed study on semi-supervised GANs for classifying cardiac disorders can be performed in the future.

- The DL models presented in this thesis are generic enough to be used for diagnosing any other cardiac disorders from single-/12-lead ECG signals. Lastly, we perceive that the lack of adequate databases has limited the research in this field. Large-scale databases featuring different cardiac diseases and advanced multimodal imaging measurements need to be created to forward research in the field.

A

Appendix: Supplementary Materials

- **Min-max normalization:** It is an amplitude scaling technique that standardizes ECG signals between 0 and 1. The general formula for the min-max normalization is as follows:

$$\mathbf{z} = \frac{\mathbf{x} - min(\mathbf{x})}{max(\mathbf{x}) - min(\mathbf{x})} \tag{A.1}$$

where $\mathbf{x} \in \mathbb{R}^T$ is the original input ECG, $\mathbf{z} \in \mathbb{R}^T$ is the normalized ECG, T denotes ECG sequence length, and *min* and *max* are the minimum and maximum values of the original input ECG.

- **Z-score normalization:** It is an amplitude scaling technique that standardizes ECG signals to have zero mean and unit variance [73]. The general formula for the Z-score normalization is as follows:

$$\mathbf{z} = \frac{\mathbf{x} - \mu}{\sigma} \tag{A.2}$$

where $\mathbf{x} \in \mathbb{R}^T$ is the original input ECG, $\mathbf{z} \in \mathbb{R}^T$ is the normalized ECG, T denotes ECG sequence length, and μ and σ are the mean and standard deviation of the original input ECG.

- **Macro-average evaluation metrics for multi-class classification:** The multi-class classification performance is evaluated using the class-wise and overall macro-average measures derived from the confusion matrix. Table A.1 shows the representation of a multi-class confusion matrix with entries u_{ab} representing the number of samples from the test set originally belonging to class a and classified as class b where $a, b \in \{1, 2, ..., C\}$. The class-wise measures include the precision (Pr), the sensitivity/recall (Re), the and the F1-score (F1). The mathematical formulation of these measures are given in Eq. A.3 where Pr_a, Re_a and the $F1_a$ are the class-wise precision, recall and F1-score for class $a, a \in \{1, 2, ..., C\}$. The overall macro-average measures, i.e., the macro-average precision, recall, and F1-score, are computed as in Eq. A.4, Eq. A.5, and Eq. A.6.

Table A.1: Multi-class confusion matrix.

		Predicted output			
		1	2	...	C
Actual output	1	u_{11}	u_{12}	...	u_{1C}
	2	u_{21}	u_{22}	...	u_{2C}
	.	.	.	...	.
	.	.	.	...	.
	.	.	.	...	.
	C	u_{C1}	u_{C2}	...	u_{CC}

$$Pr_a = \frac{u_{aa}}{\sum_{e=1}^{C} u_{ea}} \qquad Re_a = \frac{u_{aa}}{\sum_{e=1}^{C} u_{ae}} \qquad F1_a = \frac{2 \times Pr_a \times Re_a}{Pr_a + Re_a} \tag{A.3}$$

$$macro - average - precision = \frac{1}{C}\sum_{a=1}^{C} Pr_a \tag{A.4}$$

$$macro - average - recall = \frac{1}{C}\sum_{a=1}^{C} Re_a \tag{A.5}$$

$$macro - average - F1 = \frac{1}{C}\sum_{a=1}^{C} F1_a \tag{A.6}$$

- **R-peak detection-based SQI (bSQI):** In order to discard noisy ECG excerpts, we employed an R-peak or QRS detection-based signal quality index (bSQI) [8]. It compares two R-peak detectors; one is more sensitive to noise (*wqrs*) than the other (*Pan and Tompkins*). The bSQI is computed as the ratio of R-peaks detected synchronously (within an interval of 150 ms) by both detectors to all the detected R-peaks (by either detector). If both detectors agree, the clinical quality of the signals can be assumed to be high enough for automated diagnosis. Typically, excerpts with a bSQI less than 0.8 were excluded from the analysis [8].

- **Activation layer:** Activation functions generally follow the output feature maps of the convolution layers, which add non-linearity to the input features [68]. Without non-linear activations, the neural networks would perform as linear functions, thereby limiting their representation ability. The non-linear elements allow for greater flexibility and the creation of complex functions during the learning process. The rectified linear unit (ReLu) is the most popular activation function, which passes positive inputs as it is and suppresses the negative inputs to zero, thereby preventing the activation of all neurons at a time. This rectified activation reduces the computational burden of the deep neural networks, leading to faster convergence.

- **Pooling layer:** The pooling layers merge systematically similar features from the local patches of feature maps to a more robust one using max or average operations [68]. Thus, pooling layers output more robust feature maps invariant to the translation by condensing them to a smaller size. The max-pooling operation is often preferred in the convolutional layers and is performed by selecting the highest activation within the pooling window.

- **Global average pooling (GAP):** The feature maps of the last convolutional layer are transformed into a one-dimensional vector and fed to the FC layers for classification. However, the FC layers increase the computational burden, are prone to overfitting, and may hamper the generalization ability of the CNN models. Therefore, the GAP can be used in place of the FC layer, which transforms the convolutional feature maps (by averaging each) into a vector, followed by the output layer for classification. Applying GAP at the last convolutional layer can reduce the computational burden by limiting the number of trainable parameters, alleviating overfitting, and improving the generalization ability of the CNN models [73].

- **Dropout:** The Dropout layer randomly sets input units to zero with a frequency of Drop-rate at each step during training time, which prevents the network from being highly dependent on only a few neurons of the network as it may be randomly eliminated [73]. Therefore, the dropout encourages the network to spread out and assign a small quantity of weight to all the neurons. This results in shrinking the squared norm of the weights, similar to L2 regularization, thereby preventing overfitting. It should be noted that the Dropout layer is only employed during the training, and no values are dropped during inference.

- **Batch normalization (BN):** Training deep neural networks with many layers can be complicated as the distribution of each layer inputs changes for parameters update during training. This problem is also known as the "internal covariance shift," and it can be addressed by the BN technique that normalizes the outputs of the previous activation layer by subtracting the batch mean and dividing it by the batch standard deviation. Note that the batch mean and standard deviation are calculated for each output of the previous activation layer across all batch samples. The advantages of applying BN are as follows: (a) mitigates internal covariance shift, which makes training deep neural networks faster and more stable, and (b) mitigates the interdependency between layers during training, which reduces overfitting [70].

- **Identity skip connections:** To better represent the complex features, a common way used in neural networks is to stack more layers. However, the deeper networks are hard to optimize because of the vanishing gradient problem, i.e., as the gradient is back-propagated to earlier layers, repeated multiplication may make the gradient significantly small and result in slow convergence and performance degradation of the model. Introducing the identity skip connections between the adjacent layers alleviates the vanishing gradient problem by maintaining proper gradients at the initial layers, improving the deep neural network's ability to learn complex hierarchical features [153].

Table A.2: Comparison of existing ML-based approaches for MI diagnosis using ECG signals.

Reference	# Classes	# ECG-leads	Preprocessing	Input	Features	Classifier	Database	PI[†] evaluation ?	Results
Arif *et al.* [50]	2 (MI, HC)	12 (I-V6)	Baseline wander noise removal	ECG beat	Time-domain features	KNN	PTBDB	×	Acc=98.1%
Sun *et al.* [58]	2 (MI, HC)	12 (I-V6)	Baseline wander noise removal	ST-segment	ST-segment polynomial features	Multiple instance learning	PTBDB	×	Se=92.3%
Jayachandran *et al.* [48]	2 (MI, HC)	1 (II)	Baseline and high frequency noise removal	ECG beat	Entropy and energy features from wavelet subbands	Decision tree	PTBDB	×	Acc=95.0%
Sharma *et al.* [60]	2 (MI, HC)	12 (I-V6)	Baseline and high frequency noise removal	ECG segment (4 s)	Multiscale entropy and energy features	SVM-RBF	PTBDB	×	Acc=96.0%
Kumar *et al.* [62]	2 (MI, HC)	1 (II)	Baseline and high frequency noise removal	ECG beat	Entropy features from wavelet subbands	SVM-RBF	PTBDB	×	Acc=99.3%
Sadhukhan *et al.* [47]	2 (MI, HC)	3 (II, III, V2)	Normalization [0,1]	ECG beat	Phase features from Fourier transform	Logistic regression	PTBDB	×	Acc=95.6%
Acharya *et al.* [51]	2 (MI, HC)	12 (I-V6)	Baseline and high frequency noise removal	ECG beat	Entropy and nonlinear features from wavelet subbands	KNN	PTBDB	×	Acc=98.8%
Padhy *et al.* [63]	2 (MI, HC)	12 (I-V6)	HPF: 0.5 Hz	ECG beat	Energy features from multiscale HOSVD subbands	SVM-RBF	PTBDB	×	Acc=95.3%
Tripathy *et al.* [64]	2 (MI, HC)	12 (I-V6)	HPF: 0.5 Hz	ECG segment (4 s)	Entropy and statistical features from EWT subbands	SVM-RBF	PTBDB	×	Acc=99.9%
Han *et al.* [61]	2 (MI, HC)	12 (I-V6)	Baseline and high frequency noise removal	ECG beat	Entropy and statistical features from WPT subbands	SVM-RBF	PTBDB	✓	Acc=99.8%
Heden *et al.* [66]	2 (MI, HC)	12 (I-V6)	N/A	ECG beat	ST-T segment measurements	ANN	Private	×	AUC=83.4%

[†]: Patient independent (PI) evaluation. HPF: High pass filter. N/A: Not available.

Table A.3: Comparison of existing DL-based approaches for MI diagnosis using ECG signals.

Reference	# Classes	# ECG-leads	Preprocessing	Input	Features	Classifier	Database	PI† evaluation ?	Results
Acharya *et al.* [69]	2 (MI, HC)	1 (II)	Baseline wander noise removal, Z-score normalization	ECG beat	End-to-end deep CNN	Softmax	PTBDB	×	Acc=93.3%
Baloglu *et al.* [71]	2 (MI, HC)	1 (V4)	N/A	ECG beat	End-to-end deep CNN	Softmax	PTBDB	×	Acc=99.7%
Reasat *et al.* [72]	2 (IMI, HC)	3 (II, III, aVF)	Baseline and high frequency noise removal	ECG segment (3.07 s)	End-to-end shallow CNN	Softmax	PTBDB	✓	Acc=84.5%
Chen *et al.* [75]	2 (AMI, HC)	3 (V1, V2, V3)	Z-score normalization	ECG segment (1 s)	End-to-end multi-channel lightweight deep CNN	Softmax	PTBDB	×	Acc=94.2%
Liu *et al.* [73]	2 (AMI, HC)	4 (V2, V3, V5, aVL)	BPF: 0.5-100 Hz	ECG beat	End-to-end multilead CNN	Softmax	PTBDB	✓	Acc=96.0%
Cao *et al.* [74]	2 (AMI, HC)	4 (V2, V3, V5, aVL)	BPF: 0-250 Hz	ECG beat	End-to-end residual deep CNN	Softmax	PTBDB	×	Acc=96.6%
Strodthoff *et al.* [76]	2 (MI, HC)	12 (I-V6)	BPF: 0-24 Hz	ECG segment (4 s)	End-to-end residual deep CNN	Softmax	PTBDB	×	Se=93.3%
Liu *et al.* [77]	2 (MI, HC)	12 (I-V6)	N/A	ECG beat	End-to-end multiple-feature-branch CNN	Softmax	PTBDB	×	Acc=99.9%
Liu *et al.* [78]	2 (MI, HC)	12 (I-V6)	Z-score normalization	ECG beat	End-to-end multiple-feature-branch CNN+BiLSTM	Softmax	PTBDB	✓	Acc=98.7%
Han *et al.* [166]	2 (MI, HC)	12 (I-V6)	BPF: 0-100 Hz	ECG segment (4 s)	End-to-end multilead Residual deep CNN	Softmax	PTBDB	×	Acc=99.9%

†: Patient independent (PI) evaluation. BPF: Band pass filter. N/A: Not available.

Table A.4: Comparison of existing ML- and DL-based approaches for CHF diagnosis using single-lead ECG signals.

Reference	Preprocessing	Input	Features	Classifier	Database	PI† evaluation ?	Results
Melillo *et al.* [81]	N/A	RR-intervals (24 h)	Long-term HRV features, feature selection	CART	CHFDB-RRI, BIDMC-CHFDB	✓	Acc=85.4%
Shahbazi *et al.* [82]	N/A	RR-intervals (24 h)	Long-term HRV features, feature selection	KNN	CHFDB-RRI, BIDMC-CHFDB	✓	Acc=100%
Zhang *et al.* [85]	N/A	RR-intervals (30 min)	Short-term HRV features	Randon forest	BIDMC-CHFDB, NSRID	✓	Acc=91.7%
Pecchia *et al.* [84]	N/A	RR-intervals (5 min)	Short-term HRV features	Decision tree	NSRDB-RRI, CHFDB-RRI	×	Acc=93.0%
Sudarshan *et al.* [52]	Baseline wander noise removal	ECG segment (2 s)	Statistical features from DTCWT subband coefficients, feature selection	Decision tree	BIDMC-CHF, NSRDB, Fantasia	×	Acc=99.8%
Tripathy *et al.* [53]	HPF: 0.5 Hz	ECG segment (4 s)	Time-frequency entropy features	Sparse representation	BIDMC-CHFDB, NSRDB	×	Acc=98.7%
Masetic *et al.* [87]	BPF: 0.1-100 Hz	ECG segment (2.5 s)	Autoregressive parameters	Random forest	BIDMC-CHFDB, MITDB	×	Acc=100%
Acharya *et al.* [88]	Z-score normalization	ECG segment (2 s)	End-to-end deep CNN	Softmax	BIDMC-CHFDB, NSRDB, Fantasia	×	Acc=98.9%
Li *et al.* [89]	Z-score normalization	ECG segment (5 s)	End-to-end deep CNN	Softmax	Private	×	Acc=96.9%
Porumb *et al.* [49]	Z-score normalization	ECG beat	End-to-end deep CNN	Softmax	BIDMC-CHFDB, NSRDB	×	Acc=100%

†: Patient independent (PI) evaluation. BPF: Band pass filter. HPF: High pass filter. N/A: Not available.

Table A.5: Comparison of existing ML-based approaches for AF diagnosis using ECG signals.

Reference	# Classes	# ECG-leads (Device used)	Preprocessing	Input	Features	Classifier	Database	PIE† ?	Results
Tateno *et al.* [92]	2 (AF, non-AF)	1 (Holter)	N/A	RR-intervals (100)	CV	Threshold	AFDB	×	Se=86.6%
Park *et al.* [93]	2 (AF, non-AF)	1 (Holter)	BW removal	RR-intervals (30 min)	Poincare plot features	SVM-RBF	CinC2001, CinC2004	×	Se=91.4%
Sarkar *et al.* [94]	2 (AF, non-AF)	1 (Holter)	N/A	RR-intervals (2 min)	Lorenz plot features	Threshold	AFDB, NSRDB, NSRDB-RR	×	Sp=94.0%
Lian *et al.* [95]	2 (AF, non-AF)	1 (Holter)	N/A	RR-intervals (128)	Modified lorenz plot features	Threshold	AFDB, NSRDB, NSRDB-RR	×	Se=95.9%
Dash *et al.* [96]	2 (AF, non-AF)	1 (Holter)	N/A	RR-intervals (128)	TPR, RMSSD, Shannon entropy	Threshold	ADB, AFDB	×	Se=93.3%
Lee *et al.* [97]	2 (AF, NSR)	1 (Holter)	N/A	RR-intervals (64)	RMSSD and entropy features	Threshold	AFDB, NSRDB, iPhone 4S	×	Acc=98.4%
Zhou *et al.* [98]	2 (AF, non-AF)	1 (Holter)	Median filter	RR-intervals (127)	Shannon entropy features	Threshold	LTAFDB, AFDB, MITDB, NSRDB	×	Se=96.0%
Colloca *et al.* [100]	2 (AF, non-AF)	1 (Holter)	N/A	RR-intervals (300)	RR-interval variability features	SVM-RBF	AFDB, MITDB, NSRDB	✓	Se=99.0%
Carrara *et al.* [101]	3 (AF, NSR, NSR+ectopy)	1 (Holter)	g N/A	RR-intervals (10 min)	Moments, entropy, local dynamics, detrended fluctuation analysis	Random forest	Private	×	PPV=95.0%
Kennedy *et al.* [102]	2 (AF, non-AF)	1 (Holter)	Median filter	RR-intervals (30)	Sample entropy, CV, RMSSD, MAD	Random forest	LTAFDB, AFDB, ADB, NSRDB	×	Se=92.8%
Petrenas *et al.* [103]	2 (AF, non-AF)	2 (Holter)	N/A	RR-intervals and atrial activity (5 s)	RR-irregularity, P-wave absence, f-wave presence, noise level	Threshold	Private	×	Acc=88.0%
Athif *et al.* [110]	3 (AF, NSR, Others)	1 (AliveCor mobile ECG)	Lead correction	RR-intervals and atrial activity (9-60 s)	RR-irregularity, P-wave, SQI, histogram features	SVM	CinC2017	✓	F1=78.0%
Jekova *et al.* [109]	3 (AF, NSR, Others)	1 (AliveCor mobile ECG)	HPF: 1 Hz	RR-intervals and atrial activity (9-60 s)	RR-irregularity, P-wave, TQ-segment analysis	Decision tree	CinC2017	✓	F1=65.0%
Asgari *et al.* [111]	2 (AF, non-AF)	1 (Holter)	BPF: 0.5-50 Hz	ECG segment (30 s)	Entropy and peak-to-average power ratio feature from stationary WT subbands	SVM-RBF	AFDB	×	Acc=97.1%
Tripathy *et al.* [112]	2 (AF, non-AF)	1 (Holter)	BPF: 0.5-45 Hz	ECG segment (30 s)	sample entropy and center frequency features from VMD modes	DBN	AFDB	×	Acc=98.2%

†: Patient independent (PI) evaluation. BW: Baseline wander. BPF: Band pass filter. HPF: High pass filter. N/A: Not available.

Table A.6: Comparison of existing DL-based approaches for AF diagnosis using ECG signals.

Reference	# Classes	# ECG-leads (Device used)	Preprocessing	Input	Features	Classifier	Database	PIE† ?	Results
Yao *et al.* [113]	2 (AF, nonAF)	1 (Holter)	Transformations	RR-intervals (128)	End-to-end deep CNN	Softmax	AFDB, LTAFDB	×	Acc=98.1%
Faust *et al.* [114]	2 (AF, NSR)	12 (N/A)	N/A	RR-intervals (10 s)	End-to-end residual deep CNN	Softmax	Private	×	Acc=99.9%
Fang *et al.* [115]	4 (AF, NSR, Others, noise)	1 (AliveCor mobile ECG)	LPF: 30 Hz	ECG segment (9-60 s)	Poincare plot + spectrogram + end-to-end dual-channel deep CNN	Softmax	CinC2017	✓	F1=83.0%
Lai *et al.* [116]	2 (AF, NSR)	1 (Patch ECG)	HPF: 0.67 Hz	ECG segment (10 s)	Raw ECG + F-wave spectrum + RR-intervals + end-to-end deep CNN	Softmax	Private	✓	Acc=93.1%
Pourbabae *et al.* [117]	2 (AF, NSR)	1 (Holter)	Z-score	ECG segment (30 min)	End-to-end deep CNN	Softmax	CinC2001	×	Acc=85.3%
Fan *et al.* [118]	2 (AF, NSR)	1 (AliveCor mobile ECG)	Z-score	ECG segment (20 s)	End-to-end multi-scale deep CNN	Softmax	CinC2017	✓	Acc=98.1%

†: Patient independent (PI) evaluation. LPF: Low pass filter. HPF: High pass filter. N/A: Not available.

Table A.7: Comparison of existing ML- and DL-based approaches for multiple cardiac disorders classification.

Reference	# Classes	# ECG-leads	Preprocessing	Input	Features	Classifier	Database	PIE† ?	Results
Tripathy *et al.* [121]	4 (NSR, BBB, HYP, MI)	12 (I-V6)	Baseline removal	ECG segment (4 s)	Multiscale phase alternation features, feature selection	KNN	PTBDB	×	Acc=84.2%
Rahman *et al.* [122]	2 (HYP, nonHYP)	12 (I-V6)	N/A	ECG beat	Morphological and temporal features, feature selection	Random forest	Private	✓	F1=86.0%
Hannun *et al.* [15]	12 (rhythm disorders)	1 (II)	N/A	ECG segment (30 s)	End-to-end deep residual CNN	Softmax	Private	✓	F1=83.7%
Yildirim *et al.* [123]	5 (NSR and 3 rhythm disorders)	12 (I-V6)	Baseline removal	ECG segment (10 s)	End-to-end deep CNN + LSTM	Softmax	Private	✓	Acc=96.1%
Ko *et al.* [124]	2 (HYP, nonHYP)	12 (I-V6)	N/A	ECG segment (10 s)	End-to-end deep CNN	Softmax	Private	✓	AUC=96.0%
Yao *et al.* [126]	9 (NSR, AF, I-AVB, LBBB, RBBB, PAC, PVC, STE, STD)	12 (I-V6)	Z-score norm.	ECG segment (144 s)	End-to-end deep residual CNN + Bi-GRU + attention	Softmax	CPSC2018	✓	F1=81.2%
Chen *et al.* [127]	9 (NSR, AF, I-AVB, LBBB, RBBB, PAC, PVC, STE, STD)	12 (I-V6)	N/A	ECG segment (6-60 s)	End-to-end deep CNN + LSTM + attention	Softmax	CPSC2018	✓	F1=84.0%
Zhang *et al.* [128]	9 (NSR, AF, I-AVB, LBBB, RBBB, PAC, PVC, STE, STD)	12 (I-V6)	N/A	ECG segment (60 s)	End-to-end multi-lead-branch deep CNN + BiGRU + attention	Softmax	CPSC2018	✓	F1=84.3%
Wang *et al.* [129]	9 (NSR, AF, I-AVB, LBBB, RBBB, PAC, PVC, STE, STD)	12 (I-V6)	Z-score norm.	ECG segment (50 s)	End-to-end deep multiscale fusion Net	Softmax	CPSC2018	✓	F1=82.8%

†: Patient independent (PI) evaluation. N/A: Not available. I-AVB: first degree AV block. STE: ST-elevation. STD: ST-depression.

References

[1] "Cardiovascular diseases (CVDs), available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), [updated: 2021.6.11, accessed: 2022.6.14]."

[2] M. Amini, F. Zayeri, and M. Salehi, "Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017," *BMC Public Health*, vol. 21, no. 1, pp. 1–12, 2021.

[3] G. A. Roth, G. A. Mensah, C. O. Johnson, G. Addolorato, E. Ammirati, L. M. Baddour, N. C. Barengo, A. Z. Beaton, E. J. Benjamin, C. P. Benziger, *et al.*, "Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study," *Journal of the American College of Cardiology*, vol. 76, no. 25, pp. 2982–3021, 2020.

[4] "What is cardiovascular disease?, available at: https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease, [updated: 2015.7.17, accessed: 2022.6.11]."

[5] V. Kumar, A. K. Abbas, N. Fausto, and J. C. Aster, *Robbins and Cotran pathologic basis of disease, professional edition e-book*. Elsevier Health Sciences, 2014.

[6] D. S. Celermajer, C. K. Chow, E. Marijon, N. M. Anstey, and K. S. Woo, "Cardiovascular disease in the developing world: prevalences, patterns, and the potential of early disease detection," *Journal of the American College of Cardiology*, vol. 60, no. 14, pp. 1207–1216, 2012.

[7] "Diagnostic tests and procedures for heart attack, available at: https://www.heart.org/en/health-topics/heart-attack/diagnosing-a-heart-attack/diagnostic-tests-and-procedures-for-heart-attack, [updated: 2015.7.31, accessed: 2022.5.20]."

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

[9] M. A. Serhani, H. T El Kassabi, H. Ismail, and A. Nujum Navaz, "ECG monitoring systems: Review, architecture, processes, and key challenges," *Sensors*, vol. 20, no. 6, p. 1796, 2020.

[10] S. S. Al-Zaiti, Z. Faramand, K. Rjoob, D. Finlay, and R. Bond, "The role of automated 12-lead ECG interpretation in the diagnosis and risk stratification of cardiovascular disease," in *Cardiovascular and Coronary Artery Imaging*. Elsevier, 2022, pp. 45–87.

[11] A. L. Goldberger, Z. D. Goldberger, and A. Shvilkin, *Clinical electrocardiography: a simplified approach e-book*. Elsevier Health Sciences, 2017.

[12] C. Breen, R. Bond, and D. Finlay, "A clinical decision support tool to assist with the interpretation of the 12-lead electrocardiogram," *Health Informatics Journal*, vol. 25, no. 1, pp. 51–61, 2019.

[13] J. Schläpfer and H. J. Wellens, "Computer-interpreted electrocardiograms: benefits and limitations," *Journal of the American College of Cardiology*, vol. 70, no. 9, pp. 1183–1192, 2017.

[14] M. Sampson, "Ambulatory electrocardiography: indications and devices," *British Journal of Cardiac Nursing*, vol. 14, no. 3, pp. 114–121, 2019.

[15] 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.

[16] I. Markit, "The complexities of physician supply and demand: Projections from 2015 to 2030," *Assoc. Amer. Med. Colleges*, 2017.

[17] J. E. Hall and M. E. Hall, *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2020.

[18] F. Sana, E. M. Isselbacher, J. P. Singh, E. K. Heist, B. Pathik, and A. A. Armoundas, "Wearable devices for ambulatory cardiac monitoring: Jacc state-of-the-art review," *Journal of the American College of Cardiology*, vol. 75, no. 13, pp. 1582–1592, 2020.

[19] "What is atherosclerosis?, available at: https://www.nhlbi.nih.gov/health/atherosclerosis, [updated: 2022.3.22, accessed: 2022.4.26]."

[20] E. M. Antman, "Time is muscle: translation into practice," *Journal of the American College of Cardiology*, vol. 52, no. 15, pp. 1216–1221, 2008.

[21] P. T. O'gara, F. G. Kushner, D. D. Ascheim, D. E. Casey, M. K. Chung, J. A. De Lemos, S. M. Ettinger, J. C. Fang, F. M. Fesmire, B. A. Franklin, *et al.*, "2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the american college of cardiology foundation/american heart association task force on practice guidelines," *Journal of the American College of Cardiology*, vol. 61, no. 4, pp. e78–e140, 2013.

[22] H. Bulluck, D. M. Yellon, and D. J. Hausenloy, "Reducing myocardial infarct size: challenges and future opportunities," *Heart*, vol. 102, no. 5, pp. 341–348, 2016.

[23] Y. Birnbaum, J. Rankinen, H. Jneid, D. Atar, and K. Nikus, "The role of ECG in the diagnosis and risk stratification of acute coronary syndromes: an old but indispensable tool," *Current Cardiology Reports*, pp. 1–10, 2022.

[24] G. T. Lines, B. L. de Oliveira, O. Skavhaug, and M. M. Maleckar, "Simple t-wave metrics may better predict early ischemia as compared to st segment," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 6, pp. 1305–1309, 2016.

[25] A. Rosiek and K. Leksowski, "The risk factors and prevention of cardiovascular disease: the importance of electrocardiogram in the diagnosis and treatment of acute coronary syndrome," *Therapeutics and Clinical Risk Management*, vol. 12, p. 1223, 2016.

[26] "Bundle branch block: what to know, available at: https://www.webmd.com/heart-disease/bundle-branch-block, [updated: 2021.4.16, accessed: 2022.4.28]."

[27] "Bundle branch block diagnosis and treatment, available at: https://www.mayoclinic.org/diseases-conditions/bundle-branch-block/diagnosis-treatment, [updated: 2022.6.11, accessed: 2022.6.12]."

[28] "Left ventricular hypertrophy, available at: https://my.clevelandclinic.org/health/diseases/21883-left-ventricular-hypertrophy, [updated: 2021.9.20, accessed: 2022.4.13]."

[29] A. Weissler-Snir, K. Allan, K. Cunningham, K. A. Connelly, D. S. Lee, D. A. Spears, H. Rakowski, and P. Dorian, "Hypertrophic cardiomyopathy–related sudden cardiac death in young people in ontario," *Circulation*, vol. 140, no. 21, pp. 1706–1716, 2019.

[30] I. C. Van Gelder and M. E. Hemels, "The progressive nature of atrial fibrillation: a rationale for early restoration and maintenance of sinus rhythm," *Europace*, vol. 8, no. 11, pp. 943–949, 2006.

[31] P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson, D. Atar, B. Casadei, M. Castella, H.-C. Diener, H. Heidbuchel, J. Hendriks, *et al.*, "2016 esc guidelines for the management of atrial fibrillation developed in collaboration with eacts," *Kardiologia Polska (Polish Heart Journal)*, vol. 74, no. 12, pp. 1359–1469, 2016.

[32] N. R. Jones, C. J. Taylor, F. R. Hobbs, L. Bowman, and B. Casadei, "Screening for atrial fibrillation: a call for evidence," *European Heart Journal*, vol. 41, no. 10, pp. 1075–1085, 2020.

[33] B. Bozkurt, A. J. Coats, H. Tsutsui, C. M. Abdelhamid, S. Adamopoulos, N. Albert, S. D. Anker, J. Atherton, M. Böhm, J. Butler, *et al.*, "Universal definition and classification of heart failure: a report of the heart failure society of america, heart failure association of the european society of cardiology, japanese heart failure society and writing committee of the universal definition of heart failure: endorsed by the canadian heart failure society, heart failure association of india, cardiac society of australia and new zealand, and chinese heart failure association," *European Journal of Heart Failure*, vol. 23, no. 3, pp. 352–380, 2021.

[34] C. W. Yancy, M. Jessup, B. Bozkurt, J. Butler, D. E. Casey Jr, M. M. Colvin, M. H. Drazner, G. S. Filippatos, G. C. Fonarow, M. M. Givertz, *et al.*, "2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the american college of cardiology/american heart association task force on clinical practice guidelines and the heart failure society of america," *Journal of the American College of Cardiology*, vol. 70, no. 6, pp. 776–803, 2017.

[35] P. Gouda, P. Brown, B. H. Rowe, F. A. McAlister, and J. A. Ezekowitz, "Insights into the importance of the electrocardiogram in patients with acute heart failure," *European Journal of Heart Failure*, vol. 18, no. 8, pp. 1032–1040, 2016.

[36] M. J. Akhtar, M. Al-Nozha, S. Al-Harthi, and M. S. Nouh, "Electrocardiographic abnormalities in patients with heat failure," *Chest*, vol. 104, no. 2, pp. 411–414, 1993.

[37] J. Buddeke, M. L. Bots, I. van Dis, A. Liem, F. L. Visseren, and I. Vaartjes, "Trends in comorbidity in patients hospitalised for cardiovascular disease," *International Journal of Cardiology*, vol. 248, pp. 382–388, 2017.

[38] K. Rahimi, C. S. Lam, and S. Steinhubl, "Cardiovascular disease and multimorbidity: A call for interdisciplinary research and personalized cardiovascular care," p. e1002545, 2018.

[39] G. Begg, K. Willan, K. Tyndall, C. Pepper, and M. Tayebjee, "Electrocardiogram interpretation and arrhythmia management: a primary and secondary care survey," *British Journal of General Practice*, vol. 66, no. 646, pp. e291–e296, 2016.

[40] U. Satija, B. Ramkumar, and M. S. Manikandan, "A new automated signal quality-aware ECG beat classification method for unsupervised ECG diagnosis environments," *IEEE Sensors Journal*, vol. 19, no. 1, pp. 277–286, 2018.

[41] I. Jekova, V. Krasteva, R. Leber, R. Schmid, R. Twerenbold, C. Müller, T. Reichlin, and R. Abächerli, "Intersubject variability and intrasubject reproducibility of 12-lead ECG metrics: Implications for human verification," *Journal of Electrocardiology*, vol. 49, no. 6, pp. 784–789, 2016.

[42] S. Romiti, M. Vinciguerra, W. Saade, I. Anso Cortajarena, and E. Greco, "Artificial intelligence (AI) and cardiovascular diseases: an unexpected alliance," *Cardiology Research and Practice*, vol. 2020, 2020.

[43] O. Sayadi and M. B. Shamsollahi, "ECG denoising with adaptive bionic wavelet transform," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 6597–6600.

[44] M. A. Kabir and C. Shahnaz, "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains," *Biomedical Signal Processing and Control*, vol. 7, no. 5, pp. 481–489, 2012.

[45] J. Gilles, "Empirical wavelet transform," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 3999–4010, 2013.

[46] E. Prabhakararao and M. S. Manikandan, "On the use of variational mode decomposition for removal of baseline wander in ECG signals," in *2016 Twenty Second National Conference on Communication (NCC)*. IEEE, 2016, pp. 1–6.

[47] D. Sadhukhan, S. Pal, and M. Mitra, "Automated identification of myocardial infarction using harmonic phase distribution pattern of ECG data," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 10, pp. 2303–2313, 2018.

[48] E. Jayachandran, P. Joseph K, R. Acharya U, *et al.*, "Analysis of myocardial infarction using discrete wavelet transform," *Journal of Medical Systems*, vol. 34, no. 6, pp. 985–992, 2010.

[49] M. Porumb, E. Iadanza, S. Massaro, and L. Pecchia, "A convolutional neural network approach to detect congestive heart failure," *Biomedical Signal Processing and Control*, vol. 55, p. 101597, 2020.

[50] 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, no. 1, pp. 279–289, 2012.

[51] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. L. Oh, M. Adam, J. E. Koh, J. H. Tan, D. N. Ghista, R. J. Martis, C. K. Chua, *et al.*, "Automated detection and localization of myocardial infarction using electrocardiogram: a comparative study of different leads," *Knowledge-Based Systems*, vol. 99, pp. 146–156, 2016.

[52] V. K. Sudarshan, U. R. Acharya, S. L. Oh, M. Adam, J. H. Tan, C. K. Chua, K. P. Chua, and R. San Tan, "Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2 s of ECG signals," *Computers in Biology and Medicine*, vol. 83, pp. 48–58, 2017.

[53] R. K. Tripathy, M. R. Paternina, J. G. Arrieta, A. Zamora-Méndez, and G. R. Naik, "Automated detection of congestive heart failure from electrocardiogram signal using stockwell transform and hybrid classification scheme," *Computer Methods and Programs in Biomedicine*, vol. 173, pp. 53–65, 2019.

[54] 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.

[55] 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.

[56] P. Bizopoulos and D. Koutsouris, "Deep learning in cardiology," *IEEE Reviews in Biomedical Engineering*, vol. 12, pp. 168–193, 2018.

[57] S. W. Smith, B. Walsh, K. Grauer, K. Wang, J. Rapin, J. Li, W. Fennell, and P. Taboulet, "A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation," *Journal of Electrocardiology*, vol. 52, pp. 88–95, 2019.

[58] L. Sun, Y. Lu, K. Yang, and S. Li, "ECG analysis using multiple instance learning for myocardial infarction detection," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 12, pp. 3348–3356, 2012.

[59] M. Abdelazez, P. X. Quesnel, A. D. Chan, and H. Yang, "Signal quality analysis of ambulatory electrocardiograms to gate false myocardial ischemia alarms," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 6, pp. 1318–1325, 2016.

[60] 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.

[61] C. Han and L. Shi, "Automated interpretable detection of myocardial infarction fusing energy entropy and morphological features," *Computer Methods and Programs in Biomedicine*, vol. 175, pp. 9–23, 2019.

[62] M. Kumar, R. B. Pachori, and U. R. Acharya, "Automated diagnosis of myocardial infarction ECG signals using sample entropy in flexible analytic wavelet transform framework," *Entropy*, vol. 19, no. 9, p. 488, 2017.

[63] 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.

[64] R. K. Tripathy, A. Bhattacharyya, and R. B. Pachori, "A novel approach for detection of myocardial infarction from ECG signals of multiple electrodes," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4509–4517, 2019.

[65] T. Q. Le, S. T. Bukkapatnam, B. A. Benjamin, B. A. Wilkins, and R. Komanduri, "Topology and random-walk network representation of cardiac dynamics for localization of myocardial infarction," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2325–2331, 2013.

[66] B. Hedén, H. Ohlin, R. Rittner, and L. Edenbrandt, "Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks," *Circulation*, vol. 96, no. 6, pp. 1798–1802, 1997.

[67] E. Moy, M. Barrett, R. Coffey, A. L. Hines, and D. E. Newman-Toker, "Missed diagnoses of acute myocardial infarction in the emergency department: variation by patient and facility characteristics," *Diagnosis*, vol. 2, no. 1, pp. 29–40, 2015.

[68] 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.

[69] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Information Sciences*, vol. 415, pp. 190–198, 2017.

[70] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[71] U. B. Baloglu, M. Talo, O. Yildirim, R. San Tan, and U. R. Acharya, "Classification of myocardial infarction with multi-lead ecg signals and deep CNN," *Pattern Recognition Letters*, vol. 122, pp. 23–30, 2019.

[72] T. Reasat and C. Shahnaz, “Detection of inferior myocardial infarction using shallow convolutional neural networks,” in *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE, 2017, pp. 718–721.

[73] 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.

[74] 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.

[75] Y. Chen, H. Chen, Z. He, C. Yang, and Y. Cao, “Multi-channel lightweight convolution neural network for anterior myocardial infarction detection,” in *2018 IEEE SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*. IEEE, 2018, pp. 572–578.

[76] N. Strodthoff and C. Strodthoff, “Detecting and interpreting myocardial infarction using fully convolutional neural networks,” *Physiological Measurement*, vol. 40, no. 1, p. 015001, 2019.

[77] W. Liu, Q. Huang, S. Chang, H. Wang, and J. He, “Multiple-feature-branch convolutional neural network for myocardial infarction diagnosis using electrocardiogram,” *Biomedical Signal Processing and Control*, vol. 45, pp. 22–32, 2018.

[78] 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.

[79] E. E. Tripoliti, T. G. Papadopoulos, G. S. Karanasiou, K. K. Naka, and D. I. Fotiadis, “Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques,” *Computational and Structural Biotechnology Journal*, vol. 15, pp. 26–47, 2017.

[80] V. Jahmunah, S. L. Oh, J. K. E. Wei, E. J. Ciaccio, K. Chua, T. R. San, and U. R. Acharya, “Computer-aided diagnosis of congestive heart failure using ECG signals–a review,” *Physica Medica*, vol. 62, pp. 95–104, 2019.

[81] P. Melillo, N. De Luca, M. Bracale, and L. Pecchia, “Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability,” *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 727–733, 2013.

[82] F. Shahbazi and B. M. Asl, “Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability,” *Computer Methods and Programs in Biomedicine*, vol. 122, no. 2, pp. 191–198, 2015.

[83] L. Pecchia, P. Melillo, M. Sansone, and M. Bracale, “Heart rate variability in healthy people compared with patients with congestive heart failure,” in *2009 9th International Conference on Information Technology and Applications in Biomedicine*. IEEE, 2009, pp. 1–4.

[84] ——, “Discrimination power of short-term heart rate variability measures for CHF assessment,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 1, pp. 40–46, 2010.

[85] Y. Zhang, Q. Yang, W. Pang, A. Argha, P. Xu, S. Su, and D. Yao, “Congestive heart failure detection via short-time electrocardiographic monitoring for fast reference advice in urgent medical conditions,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 2256–2259.

[86] A. Bansal and R. Joshi, “Portable out-of-hospital electrocardiography: A review of current technologies,” *Journal of Arrhythmia*, vol. 34, no. 2, pp. 129–138, 2018.

[87] Z. Masetic and A. Subasi, “Congestive heart failure detection using random forest classifier,” *Computer Methods and Programs in Biomedicine*, vol. 130, pp. 54–64, 2016.

[88] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, and R. S. Tan, “Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals,” *Applied Intelligence*, vol. 49, no. 1, pp. 16–27, 2019.

[89] D. Li, X. Li, J. Zhao, and X. Bai, “Automatic staging model of heart failure based on deep learning,” *Biomedical Signal Processing and Control*, vol. 52, pp. 77–83, 2019.

[90] M. Butkuvienė, A. Petrėnas, A. Sološenko, A. Martín-Yebra, V. Marozas, and L. Sörnmo, “Considerations on performance evaluation of atrial fibrillation detectors,” *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 11, pp. 3250–3260, 2021.

[91] G. Moody, “A new method for detecting atrial fibrillation using RR intervals,” *Computers in Cardiology*, pp. 227–230, 1983.

[92] K. Tateno and L. Glass, “Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and δRR intervals,” *Medical and Biological Engineering and Computing*, vol. 39, no. 6, pp. 664–671, 2001.

[93] J. Park, S. Lee, and M. Jeon, “Atrial fibrillation detection by heart rate variability in Poincare plot,” *Biomedical Engineering Online*, vol. 8, no. 1, pp. 1–12, 2009.

[94] S. Sarkar, D. Ritscher, and R. Mehra, “A detector for a chronic implantable atrial tachyarrhythmia monitor,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1219–1224, 2008.

[95] J. Lian, L. Wang, and D. Muessig, “A simple method to detect atrial fibrillation using RR intervals,” *The American Journal of Cardiology*, vol. 107, no. 10, pp. 1494–1497, 2011.

[96] S. Dash, K. Chon, S. Lu, and E. Raeder, “Automatic real time detection of atrial fibrillation,” *Annals of Biomedical Engineering*, vol. 37, no. 9, pp. 1701–1709, 2009.

[97] J. Lee, B. A. Reyes, D. D. McManus, O. Maitas, and K. H. Chon, “Atrial fibrillation detection using an iphone 4S,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 203–206, 2012.

[98] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, and Y. Zhang, “Automatic online detection of atrial fibrillation based on symbolic dynamics and shannon entropy,” *Biomedical Engineering Online*, vol. 13, no. 1, pp. 1–18, 2014.

[99] R. Alcaraz, D. Abásolo, R. Hornero, and J. J. Rieta, “Optimal parameters study for sample entropy-based atrial fibrillation organization analysis,” *Computer Methods and Programs in Biomedicine*, vol. 99, no. 1, pp. 124–132, 2010.

[100] R. Colloca, A. E. Johnson, L. Mainardi, and G. D. Clifford, “A support vector machine approach for reliable detection of atrial fibrillation events,” in *Computing in Cardiology 2013*. IEEE, 2013, pp. 1047–1050.

[101] M. Carrara, L. Carozzi, T. J. Moss, M. De Pasquale, S. Cerutti, M. Ferrario, D. E. Lake, and J. R. Moorman, “Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy,” *Physiological Measurement*, vol. 36, no. 9, p. 1873, 2015.

[102] A. Kennedy, D. D. Finlay, D. Guldenring, R. R. Bond, K. Moran, and J. McLaughlin, “Automated detection of atrial fibrillation using RR intervals and multivariate-based classification,” *Journal of Electrocardiology*, vol. 49, no. 6, pp. 871–876, 2016.

[103] A. Petrėnas, V. Marozas, and L. Sörnmo, “Low-complexity detection of atrial fibrillation in continuous long-term monitoring,” *Computers in Biology and Medicine*, vol. 65, pp. 184–191, 2015.

[104] S. Ladavich and B. Ghoraani, “Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity,” *Biomedical Signal Processing and Control*, vol. 18, pp. 274–281, 2015.

[105] R. Alcaraz and J. J. Rieta, “Adaptive singular value cancelation of ventricular activity in single-lead atrial fibrillation electrocardiograms,” *Physiological Measurement*, vol. 29, no. 12, p. 1351, 2008.

[106] J. P. Ugarte, A. Orozco-Duque, C. Tobón, V. Kremen, D. Novak, J. Saiz, T. Oesterlein, C. Schmitt, A. Luik, and J. Bustamante, “Dynamic approximate entropy electroanatomic maps detect rotors in a simulated atrial fibrillation model,” *PloS One*, vol. 9, no. 12, p. e114577, 2014.

[107] T. Lankveld, S. Zeemering, D. Scherr, P. Kuklik, B. A. Hoffmann, S. Willems, B. Pieske, M. Haïssaguerre, P. Jaïs, H. J. Crijns, *et al.*, “Atrial fibrillation complexity parameters derived from surface ECGs predict procedural outcome and long-term follow-up of stepwise catheter ablation for atrial fibrillation,” *Circulation: Arrhythmia and Electrophysiology*, vol. 9, no. 2, p. e003354, 2016.

[108] A. Petrėnas, L. Sörnmo, A. Lukoševičius, and V. Marozas, “Detection of occult paroxysmal atrial fibrillation,” *Medical & Biological Engineering & Computing*, vol. 53, no. 4, pp. 287–297, 2015.

[109] I. I. Jekova, T. V. Stoyanov, and I. A. Dotsinsky, “Arrhythmia classification via time and frequency domain analyses of ventricular and atrial contractions,” in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[110] M. Athif, P. C. Yasawardene, and C. Daluwatte, "Detecting atrial fibrillation from short single lead ECGs using statistical and morphological features," *Physiological measurement*, vol. 39, no. 6, p. 064002, 2018.

[111] S. Asgari, A. Mehrnia, and M. Moussavi, "Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine," *Computers in Biology and Medicine*, vol. 60, pp. 132–142, 2015.

[112] R. Tripathy, M. R. A. Paternina, J. G. Arrieta, and P. Pattanaik, "Automated detection of atrial fibrillation ECG signals using two stage VMD and atrial fibrillation diagnosis index," *Journal of Mechanics in Medicine and Biology*, vol. 17, no. 07, p. 1740044, 2017.

[113] Z. Yao, Z. Zhu, and Y. Chen, "Atrial fibrillation detection by multi-scale convolutional neural networks," in *2017 20th International Conference on Information Fusion (Fusion)*. IEEE, 2017, pp. 1–6.

[114] O. Faust, M. Kareem, A. Ali, E. J. Ciaccio, and U. R. Acharya, "Automated arrhythmia detection based on RR intervals," *Diagnostics*, vol. 11, no. 8, p. 1446, 2021.

[115] B. Fang, J. Chen, Y. Liu, W. Wang, K. Wang, A. K. Singh, and Z. Lv, "Dual-channel neural network for atrial fibrillation detection from a single lead ECG wave," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[116] D. Lai, Y. Bu, Y. Su, X. Zhang, and C.-S. Ma, "Non-standardized patch-based ECG lead together with deep learning based algorithm for automatic screening of atrial fibrillation," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 6, pp. 1569–1578, 2020.

[117] 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, 2018.

[118] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1744–1753, 2018.

[119] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ECG analysis: Benchmarks and insights from PTB-XL," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1519–1528, 2020.

[120] M. Yahalom, N. Roguin, K. Suleiman, and Y. Turgeman, "Clinical significance of conditions presenting with ecg changes mimicking acute myocardial infarction," *International Journal of Angiology*, vol. 22, no. 02, pp. 115–122, 2013.

[121] R. Tripathy and S. Dandapat, "Detection of cardiac abnormalities from multilead ECG using multiscale phase alternation features," *Journal of Medical Systems*, vol. 40, no. 6, pp. 1–9, 2016.

[122] Q. A. Rahman, L. G. Tereshchenko, M. Kongkatong, T. Abraham, M. R. Abraham, and H. Shatkay, "Utilizing ECG-based heartbeat classification for hypertrophic cardiomyopathy identification," *IEEE Transactions on Nanobioscience*, vol. 14, no. 5, pp. 505–512, 2015.

[123] O. Yildirim, M. Talo, E. J. Ciaccio, R. San Tan, and U. R. Acharya, "Accurate deep neural network model to detect cardiac arrhythmia on more than 10,000 individual subject ECG records," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105740, 2020.

[124] W.-Y. Ko, K. C. Siontis, Z. I. Attia, R. E. Carter, S. Kapa, S. R. Ommen, S. J. Demuth, M. J. Ackerman, B. J. Gersh, A. M. Arruda-Olson, *et al.*, "Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram," *Journal of the American College of Cardiology*, vol. 75, no. 7, pp. 722–733, 2020.

[125] 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.

[126] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, "Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional neural network," *Information Fusion*, vol. 53, pp. 174–182, 2020.

[127] T.-M. Chen, C.-H. Huang, E. S. Shih, Y.-F. Hu, and M.-J. Hwang, "Detection and classification of cardiac arrhythmias by a challenge-best deep learning neural network model," *Iscience*, vol. 23, no. 3, p. 100886, 2020.

[128] J. Zhang, D. Liang, A. Liu, M. Gao, X. Chen, X. Zhang, and X. Chen, "MLBF-Net: a multi-lead-branch fusion network for multi-class arrhythmia classification using 12-lead ECG," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 9, pp. 1–11, 2021.

[129] R. Wang, J. Fan, and Y. Li, “Deep multi-scale fusion neural network for multi-class arrhythmia detection,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 9, pp. 2461–2472, 2020.

[130] Y. Li, Z. Zhang, F. Zhou, Y. Xing, J. Li, and C. Liu, “Multi-label classification of arrhythmia for long-term electrocardiogram signals with feature learning,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[131] D. Jia, W. Zhao, Z. Li, C. Yan, H. Wang, J. Hu, and J. Fang, “An ensemble neural network for multi-label classification of electrocardiogram,” in *Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting*. Springer, 2019, pp. 20–27.

[132] M. Ganeshkumar, V. Ravi, V. Sowmya, E. Gopalakrishnan, and K. Soman, “Explainable deep learning-based approach for multilabel classification of electrocardiogram,” *IEEE Transactions on Engineering Management*, 2021.

[133] Y. Jin, J. Liu, Y. Liu, C. Qin, Z. Li, D. Xiao, L. Zhao, and C. Liu, “A novel interpretable method based on dual level attentional deep neural network for actual multi label arrhythmia detection,” *IEEE Transactions on Instrumentation and Measurement*, 2021.

[134] G. Boriani, I. Diemberger, M. Ziacchi, C. Valzania, B. Gardini, P. Cimaglia, C. Martignani, and M. Biffi, “AF burden is important–fact or fiction?” *International Journal of Clinical Practice*, vol. 68, no. 4, pp. 444–452, 2014.

[135] G. Boriani and D. Pettorelli, “Atrial fibrillation burden and atrial fibrillation type: clinical significance and impact on the risk of stroke and decision making for long-term anticoagulation,” *Vascular Pharmacology*, vol. 83, pp. 26–35, 2016.

[136] L. Y. Chen, M. K. Chung, L. A. Allen, M. Ezekowitz, K. L. Furie, P. McCabe, P. A. Noseworthy, M. V. Perez, and M. P. Turakhia, “Atrial fibrillation burden: moving beyond atrial fibrillation as a binary entity: a scientific statement from the American Heart Association,” *Circulation*, vol. 137, no. 20, pp. e623–e644, 2018.

[137] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, “Adaptive mixtures of local experts,” *Neural Computation*, vol. 3, no. 1, pp. 79–87, 1991.

[138] S. Mavandadi, S. Feng, F. Yu, S. Dimitrov, K. Nielsen-Saines, W. R. Prescott, and A. Ozcan, “A mathematical framework for combining decisions of multiple experts toward accurate and remote diagnosis of malaria using tele-microscopy,” 2012.

[139] A. Graves and N. Jaitly, “Towards end-to-end speech recognition with recurrent neural networks,” in *International Conference on Machine Learning*. PMLR, 2014, pp. 1764–1772.

[140] Y. Yu, X. Si, C. Hu, and J. Zhang, “A review of recurrent neural networks: LSTM cells and network architectures,” *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.

[141] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *arXiv preprint arXiv:1409.0473*, 2014.

[142] J. M. McCabe, E. J. Armstrong, I. Ku, A. Kulkarni, K. S. Hoffmayer, P. D. Bhave, S. W. Waldo, P. Hsue, J. C. Stein, G. M. Marcus, *et al.*, “Physician accuracy in interpreting potential ST-segment elevation myocardial infarction electrocardiograms,” *Journal of the American Heart Association*, vol. 2, no. 5, p. e000268, 2013.

[143] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, “Show, attend and tell: Neural image caption generation with visual attention,” in *International Conference on Machine Learning*. PMLR, 2015, pp. 2048–2057.

[144] M. Chen, X. He, J. Yang, and H. Zhang, “3-D convolutional recurrent neural networks with attention model for speech emotion recognition,” *IEEE Signal Processing Letters*, vol. 25, no. 10, pp. 1440–1444, 2018.

[145] D. Zhang, L. Yao, K. Chen, and J. Monaghan, “A convolutional recurrent attention model for subject-independent EEG signal analysis,” *IEEE Signal Processing Letters*, vol. 26, no. 5, pp. 715–719, 2019.

[146] S. Brownlee, K. Chalkidou, J. Doust, A. G. Elshaug, P. Glasziou, I. Heath, S. Nagpal, V. Saini, D. Srivastava, K. Chalmers, *et al.*, “Evidence for overuse of medical services around the world,” *The Lancet*, vol. 390, no. 10090, pp. 156–168, 2017.

[147] 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.

[148] 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.

[149] E. Prabhakararao and S. Dandapat, "Automatic quality estimation of 12-lead ECG for remote healthcare monitoring systems," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE, 2018, pp. 554–559.

[150] S. S. Xu, M.-W. Mak, and C.-C. Cheung, "Towards end-to-end ECG classification with raw signal extraction and deep neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1574–1584, 2018.

[151] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012.

[152] S. S. Anand, S. Islam, A. Rosengren, M. G. Franzosi, K. Steyn, A. H. Yusufali, M. Keltai, R. Diaz, S. Rangarajan, and S. Yusuf, "Risk factors for myocardial infarction in women and men: insights from the INTERHEART study," *European Heart Journal*, vol. 29, no. 7, pp. 932–940, 2008.

[153] 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.

[154] J. Kornej, C. S. Börschel, E. J. Benjamin, and R. B. Schnabel, "Epidemiology of atrial fibrillation in the 21st century: novel methods and new insights," *Circulation Research*, vol. 127, no. 1, pp. 4–20, 2020.

[155] A. S. Go, K. Reynolds, J. Yang, N. Gupta, J. Lenane, S. H. Sung, T. N. Harrison, T. I. Liu, and M. D. Solomon, "Association of burden of atrial fibrillation with risk of ischemic stroke in adults with paroxysmal atrial fibrillation: the KP-RHYTHM study," *JAMA Cardiology*, vol. 3, no. 7, pp. 601–608, 2018.

[156] A. Chocron, J. Oster, S. Biton, F. Mandel, M. Elbaz, Y. Y. Zeevi, and J. A. Behar, "Remote atrial fibrillation burden estimation using deep recurrent neural network," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 8, pp. 2447–2455, 2020.

[157] K. Ochiai, S. Takahashi, and Y. Fukazawa, "Arrhythmia detection from 2-lead ECG using convolutional denoising autoencoders," in *Proc. KDD*, 2018, pp. 1–7.

[158] A. Amaraneni, V. Chippa, and A. C. Rettew, "Anticoagulation safety," in *StatPearls [Internet]*. StatPearls Publishing, 2021.

[159] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[160] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, and A. Elmaghraby, "Ensemble deep learning models for heart disease classification: a case study from mexico," *Information*, vol. 11, no. 4, p. 207, 2020.

[161] Y. Liu and X. Yao, "Simultaneous training of negatively correlated neural networks in an ensemble," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 6, pp. 716–725, 1999.

[162] 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, pp. 1–15, 2020.

[163] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, A. Johnson, and R. G. Mark, "AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[164] C. O. Freitas, J. M. d. Carvalho, J. Oliveira, S. B. Aires, and R. Sabourin, "Confusion matrix disagreement for multiple classifiers," in *Iberoamerican Congress on Pattern Recognition*. Springer, 2007, pp. 387–396.

[165] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81–90, 2017.

[166] 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.

[167] 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.

[168] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[169] Y. Gao, H. Wang, and Z. Liu, "A novel approach for atrial fibrillation signal identification based on temporal attention mechanism," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 316–319.

[170] A. Graves, "Long short-term memory," *Supervised sequence labelling with recurrent neural networks*, pp. 37–45, 2012.

[171] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, 2013.

[172] A. P. Gorgels, D. Engelen, and H. J. Wellens, "Lead aVR, a mostly ignored but very valuable lead in clinical electrocardiography," pp. 1355–1356, 2001.

[173] J. Abbasi, "The COVID heart—one year after SARS-CoV-2 infection, patients have an array of increased cardiovascular risks," *JAMA*, vol. 327, no. 12, pp. 1113–1114, 2022.

[174] E. Prabhakararao and S. Dandapat, "Automated detection of posterior myocardial infarction from VCG signals using stationary wavelet transform based features," *IEEE Sensors Letters*, vol. 4, no. 6, pp. 1–4, 2020.

[175] J. Poldervaart, M. Langedijk, B. Backus, I. Dekker, A. Six, P. Doevendans, A. Hoes, and J. Reitsma, "Comparison of the GRACE, HEART and TIMI score to predict major adverse cardiac events in chest pain patients at the emergency department," *International Journal of Cardiology*, vol. 227, pp. 656–661, 2017.

[176] L. Saba, "Advanced imaging in the diagnosis of cardiovascular diseases: the "ongoing" future," *Cardiovascular Diagnosis and Therapy*, vol. 10, no. 4, p. 915, 2020.

[177] G. R. Marzouka, H. Rivner, V. Mehta, J. Lopez, I. Vaz, F. Tang, H. Ishwaran, and J. J. Goldberger, "The CHA2DS2-VASc score for risk stratification of stroke in heart failure with-vs-without atrial fibrillation," *The American Journal of Cardiology*, vol. 155, pp. 72–77, 2021.

[178] A. Madani, J. R. Ong, A. Tibrewal, and M. R. Mofrad, "Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease," *NPJ Digital Medicine*, vol. 1, no. 1, pp. 1–11, 2018.

[179] V. Das, S. Dandapat, and P. K. Bora, "A data-efficient approach for automated classification of OCT images using generative adversarial network," *IEEE Sensors Letters*, vol. 4, no. 1, pp. 1–4, 2020.

List of Publications

Journal Publications:

1. Eedara Prabhakararao and Samarendra Dandapat, “Automated Detection of Congestive Heart Failure from ECG Signals using Deep Residual Recurrent Neural Network,” **IEEE Transactions on Systems, Man, and Cybernetics: Systems,** Dec. 2022, DOI: 10.1109/TSMC.2022.3221843.

2. Eedara Prabhakararao and Samarendra 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, Dec. 2022.

3. Eedara Prabhakararao and Samarendra Dandapat, “Multi-Scale Convolutional Neural Network Ensemble for Multi-Class Arrhythmia Classification,” **IEEE Journal of Biomedical and Health Informatics,** vol. 26, no. 8, pp. 3802-3812, Dec. 2021.

4. Eedara Prabhakararao and Samarendra Dandapat, “Attentive RNN-Based Network to Fuse 12-Lead ECG and Clinical Features for Improved Myocardial Infarction Diagnosis,” **IEEE Signal Processing Letters,** vol. 27, pp. 2029-2033, 2020.

5. Eedara Prabhakararao and Samarendra 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.

6. Eedara Prabhakararao and Samarendra Dandapat, “Automated Detection of Posterior Myocardial Infarction from VCG Signals using Stationary Wavelet Transform based Features,” **IEEE Sensors Letters,** vol. 4, no. 6, pp. 1-4, 2020.

7. Eedara Prabhakararao and Samarendra Dandapat, “Multi-Label ECG classification using Attention based Temporal Convolutional Neural Network,” **IEEE Transactions on Instrumentation and Measurement,** [Status: Under review].

Conference Publications:

1. Eedara Prabhakararao and Samarendra Dandapat, “Multiscale Convolutional Neural Network for Detecting Paroxysmal Atrial Fibrillation from Single-Lead ECG,” in **Proceedings of the IEEE International Conference on Applied Signal Processing (ASPCON), Kolkata**, pp. 339-343, 2020.

2. Eedara Prabhakararao and Samarendra Dandapat, "A weighted SVM based Approach for Automatic Detection of Posterior Myocardial Infarction using VCG Signals," in **Proceedings of the IEEE National Conference on Communications (NCC), IISc Bangalore**, pp. 1-6, 2020.

3. Eedara Prabhakararao and Samarendra Dandapat, "Automatic Quality Estimation of 12-lead ECG for Remote Healthcare Monitoring Systems," in **Proceedings of the IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Malaysia**, pp. 554-559, 2018.