AN ELECTROCARDIOGRAM BASED SECURE PERSON ADAPTIVE CARDIOVASCULAR DISEASE DIAGNOSIS SYSTEM

DEBASISH JYOTISHI

AN ELECTROCARDIOGRAM BASED SECURE PERSON ADAPTIVE CARDIOVASCULAR DISEASE DIAGNOSIS SYSTEM

A

Thesis submitted

for the award of the degree of

DOCTOR OF PHILOSOPHY

By

DEBASISH JYOTISHI

DEPARTMENT OF ELECTRONICS AND ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

GUWAHATI - 781 039, ASSAM, INDIA

JUNE 2024

Certificate

This is to certify that the thesis entitled **"An Electrocardiogram Based Secure Person Adaptive Cardiovascular Disease Diagnosis System"**, submitted by **Debasish Jyotishi**, Roll No. **186102006**, 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, in my opinion, has reached the standard needed for the submission. The results embodied in this thesis have not been submitted to any other university or institute for the award of any degree or diploma.

Dated:

Place: IIT Guwahati

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

Dedicated

To

Acknowledgements

I am deeply grateful to Prof. Samarendra Dandapat for his unwavering support, invaluable guidance, nurturing approach, and genuine concern throughout my doctoral journey. His scientific insights have significantly influenced and enriched this work. The constant motivation and conducive research environment he provided were instrumental in conducting this research work. Beyond research, I am thankful for the enriching life experiences gained during this period. It has been an absolute privilege and pleasure to work under his mentorship.

I extend my heartfelt gratitude to my doctoral committee members, Prof. Rohit Sinha, Prof. P. K. Bora, and Dr. Salil Kashyap, for their encouragement and invaluable suggestions on my work. Their deep insights, and constructive criticism have been instrumental in shaping my research into its current form. I am also thankful to the faculty members of the Department of Electronics and Electrical Engineering (EEE) whose guidance and expertise have enriched my technical knowledge and understanding. I would like to thank Prof. V. Ramakrishnan for his help and support. My appreciation goes to the staff of the EEE Department for their consistent help and support throughout my thesis work. Furthermore, I am thankful to all the staff members of IIT Guwahati who have influenced my life during the PhD journey. A special acknowledgment goes to Mr. Bhriguraj Borah, technical staff of the Department of Computer Science and Engineering (CSE), for his diligent maintenance and timely provision of computational facilities.

I would also like to express my gratitude to Sujata Ma'am, whose caring presence made me feel at home despite being far away from home.

I would like to express my heartfelt thanks to my seniors, Dr. Suman Deb, Dr. Tilendra Choudhary, Dr. Sumit Dutta, Dr. Bikash Sah for their invaluable help and unwavering support.

I am deeply grateful to my friends, Dr. Sibasis Sahoo, Himashree Kalita, Dr. Samarjeet Das, Pharvesh Salman Choudhury, Mousumi Das, Atanu Purkayastha, Sumit Singha, Moirangthem James Singh, Yengkhom Omesh Singh, Dr. Alex P. Kamson, Ato Kapfo, Dr. Vineeta Das, Dr. E. Prabhakararao, Akriti Jaiswal, and Mane Pooja, for their generous assistance throughout my research work. Special thanks to Himashree Kalita for being a constant source of support through both highs and lows. My sincere gratitude extends to Dr. L. N. Sharma and all the research scholars in the Electro-Medical Speech Technology (EMST) and Signal Informatics Lab for their valuable support

during my research.

I am deeply thankful to my family members for their grace and unwavering support. I acknowledge the profound impact of my mother's sacrifices, love, and care in shaping my life. I am grateful for my father's support and scientific temperament, which has deeply influenced my thought process and understanding. I also appreciate the love of my sister, brother-in-law, and nieces. I acknowledge the invaluable contributions of all my past teachers in molding my academic knowledge.

Finally, I acknowledge everything that is and that is not; which has been a powerful force of presence during all these days.

Shambho

Debasish Jyotishi

Abstract

The electrocardiogram (ECG) signal, which records the heart's electrical activity, encapsulates valuable diagnostic and biometric information. ECG signal is the primary non-invasive tool used by the cardiologists for diagnosing various cardiovascular diseases (CVDs). CVDs are one of the major cause of premature deaths worldwide. Early detection and diagnosis of CVDs are essential for effective treatment and cure of cardiac diseases. As a result, there have been significant technological advancements, enabling one-time ECG recordings, continuous heart monitoring, ambulatory recordings, and remote monitoring. This has resulted in abundance of ECG data for interpretation by experienced cardiologists. However, the manual interpretation is time consuming, prone to human errors and depends on the availability of expert cardiologists. Consequently, automated CVD diagnosis systems have been developed to assist cardiologists. However, the major challenge towards developing a robust CVD diagnosis system is the inter-individual variation of the morphological characteristics of the ECG signal. This necessitates the development of person adaptive CVD diagnosis system, which requires the understanding of the person specific information present in the ECG signal. The person specific information can be further extended to develop ECG based biometric system, offering a promising solution to the security and privacy concern for wearable healthcare devices and sensitive medical data. The ECG signal is also a suitable biometric modality for patient identification in healthcare system, which is a crucial task for patient safety.

The objective of the thesis is to design automated models for learning deep temporal and spatio-temporal representation from multi-lead ECG signal, with the overarching goal of developing robust biometric and automated CVD diagnosis system. Furthermore, a novel method is devised to effectively use the learned person-specific representations to develop a person-adaptive CVD diagnosis system.

The ECG signal is a cyclostationary signal, with each ECG beat representing the depolar-

isation and relaxation of heart chambers. Variations in the physiological and geometrical characteristics of individuals' hearts lead to distinct morphological characteristics in ECG signals. These changes manifest in the wave shapes and temporal dynamics of the ECG signal. In the first work, we developed an ECG based person identification and verification system by learning the underlying temporal representation of the ECG signal. We designed a long short-term memory (LSTM) based framework for learning the intra-beat and inter-beat variations present in the ECG signal. This is achieved by training the LSTM network with smaller segments of ECG signals as input, which are extracted by sliding a rectangular window. A major advantage of the proposed method is it doesn't require any fiducial point detection. Experimental results suggest that the LSTM model can capture the intrabeat variations better for smaller ECG segments, leading to better identification performance. However, we observed that the LSTM model suffers modeling the long-term temporal dependencies in the ECG signal and it lacks modeling multi-scale temporal representation. Motivated by this we designed a novel attention based hierarchical LSTM (HLSTM) model to learn the biometric representation. HLSTM model learns the temporal variation of the ECG signal in different abstractions. This addresses the long term temporal dependency issue of the LSTM network in our application. The attention mechanism of the model learns to capture the ECG complexes that have more biometric information corresponding to each person. These ECG complexes are given more weight to learn better biometric representation. Empirical findings demonstrate promising results, showing substantial performance enhancements achieved through the utilization of multi-scale temporal information.

In the second work, we proposed a novel biometric framework by capturing both the local morphological representation and multi-scale temporal dynamics of ECG signals. We introduced a novel multi-scale temporal dynamics learning network (MSTDLNet) for the acquisition of robust biometric features. The MSTDLNet architecture comprises a multi-scale enhanced morphological representation learning (MSE-MRL) module and two layers of LSTM network. The MSE-MRL module is designed to learn local multi-scale morphological representations while emphasizing specific ECG morphologies to enrich the biometric features. The LSTM networks are innovatively integrated to the MSE-MRL

module to learn the multi-scale temporal representation leading to robust biometric representation. Experimental results demonstrate the MSTDLNet model's superior capability to learn robust multi-scale temporal representation, which results in state-of-the-art performance in biometric identification.

In the third work, we proposed an automated CVD diagnosis system using multi-lead ECG signal. Designing an automated system for diagnosing multiple cardiac abnormalities is a challenging task due to the tenuous morphological variation of the ECG signal across different cardiac diseases. In this work, we developed an attentive spatio-temporal learning network (ASTLNet) that can learn better diagnostic representation by exploiting the concurrent spatio-temporal variation of the multilead ECG signal. The ASTLNet consists of two modules: spatio-temporal representation learning (STRL) module and attentive spatio-temporal aggregation (ASTA) module. The STRL module is designed to learn the multiscale spatio-temporal representation, and the ASTA module is designed to aggregate the learned representation. Experiments conducted on publicly available datasets demonstrate that the proposed model can effectively learn the spatio-temporal variation of the ECG signal leading to improved diagnostic outcome.

In the last work, we designed a person-adaptive CVD diagnosis system by introducing an attention based memory module and conditional normalization. The person adaptive CVD diagnosis sytem is designed in an modular structure which allows for all types of global CVD diagnosis models. The person specific ECG information are embedded into the diagnostic models through the conditional normalization facilitated by a conditioning network. The parameters of the conditioning network are controlled by a memory module which encapsulates the person specific information by using the ECG-based biometric representation (iECG vector). An auxiliary attention network generates a memory vector for a new test subject in an unsupervised manner that leverages the person-specific information encapsulated in the memory module. Experimental results show that the person-adaptive CVD diagnosis systems improve the diagnostic performance significantly compared to a global CVD diagnosis system.

Keywords: Electrocardiogram, Cardiovascular Disease, Biometrics, person-adaptive CVD diagnosis system, Long short term memory (LSTM) network, Representation learning.

Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
AFIB	Atrial Fibrillation
AFLT	Atrial Flutter
AV	Atrioventricular
AVNRT	Atrioventricular Nodal Reentrant Tachycardia
AVRT	Atrioventricular Reentrant Tachycardia
BBB	Bundle Branch Block
BIGU	Bigeminal Pattern (unknown origin, SV or Ventricular)
CAD	Coronary Artery Disease
CHF	Congestive Heart Failure
CNN	Convolutional Neural Network
CVD	Cardio Vascular Disease
DCT	Discrete Cosine Transform
DL	Deep Learning
DNN	Deep Neural Network
DCT	Discrete Cosine Transform
DL	Deep Learning
ECG	Electrocardiogram
EER	Equal Error Rate
FAR	False Acceptance Rate
FRR	False Rejection Rate
HMM	Hidden Markov Models
IoT	Internet of Things

LA	Left Atrium
LAO	Left Atrial Overload
LBBB	Left Bundle Branch Block
LDA	Linear Discriminant Analysis
LV	Left Ventricle
LVH	Left Ventricular Hypertrophy
MI	Myocardial Infarction
PACE	Normal Functioning Artificial Pacemaker
PCA	Principal Component Analysis
PSVT	Paroxysmal Supraventricular Tachycardia
PVCs	Premature Ventricular Complexes
RA	Right Atrium
RAO	Right Atrial Overload
RBBB	Right Bundle Branch Block
RNN	Recurrent Neural Network
RV	Left Ventricle
RVH	Right Ventricular Hypertrophy
SA	Sino-Atrial
SARRH	Sinus Arrhythmia
SBRAD	Sinus Bradycardia
SR	Sinus Rhythm
STACH	Sinus Tachycardia
SVD	Singular Value Decomposition
SNR	Signal to Noise Ratio
SVARR	Supraventricular arrhythmia
SVTAC	Supraventricular Tachycardia
TRIGU	Trigeminal Pattern (unknown origin, SV or Ventricular)
VCG	Vectorcardiogram
VER	Ventricular Escape Rhythm
VFIB	Ventricular Fibrillation

VTAC Ventricular Tachycardia

1

Introduction

Contents

1. Introduction

Cardiovascular diseases (CVDs) are the leading cause of deaths and disability around the globe [1, 2]. CVDs are responsible for an estimated 17.9 million deaths in the year 2019, which accounts for 32% of global deaths [2]. CVDs contribute approximately 38% of the total premature deaths worldwide [2]. Among the wide spectrum of CVDs, heart diseases, including coronary artery disease (CAD) and cardiac arrhythmias, stand out as significant contributors. Early and accurate diagnosis of cardiac abnormalities is pivotal for guiding timely interventions and optimizing patient outcomes. Electrocardiogram (ECG), which records the electrical activity of the heart, is a primary clinical non-invasive method for diagnosing majority of cardiac abnormalities. In clinical practice, ECG signals are typically recorded in two primary setups,i.e., Holter ECG recording and 12-lead ECG recording. The multi-lead ECG signal views the electrical activity of the heart from different direction. The multi-lead ECG signal presents a spatio-temporal variation of the heart's electrical activity, which presents vital clinical information for diagnosing cardiac ailments. However, the interpretation of large amount of multi-lead ECG signal generated through continuous monitoring is an arduous task and prone to errors. It is also subject to the availability of expert cardiologist. Hence, there exist a pressing need for the development of automated CVD diagnosis systems to expedite informed clinical decision making process.

Developing a robust automated CVD diagnosis system is a challenging task due to the variation in the morphological characteristics of the ECG signal across diverse categories of subjects, termed as inter-individual variability [3–5]. The inter-individual variability, caused due to many-fold generative factors, hampers the diagnostic performance of the global CVD diagnosis systems trained on ECG data taken from a large pool of subjects [3, 6, 7]. This is because the trained global models are unable to generalize to data from a new test subject, characterised by a different underlying distribution compared to the training data. Consequently, this challenge has spurred the development of personalized diagnostic systems—a burgeoning domain within the broader framework of precision medicine [8, 9], which has garnered attention in recent times.

The personalized CVD diagnostic system is increasingly relevant today due to the extensive use of wearable health devices and remote healthcare for better health monitoring. Advances in the field of communication and internet of things (IoT) have shifted healthcare system towards artificial intelligence based (AI) based healthcare system. This has raised the concern for the security and privacy of sensitive medical data recorded both in the personal wearable devices as well as in hospital

setup. The ECG signal can be used as a biometric modality for information security in healthcare services. ECG signal has distinct morphological pattern corresponding to each person and it is robust against spoof attacks due to its inherent liveliness information and less exposure to covert acquisition. Except for the cardiovascular disease cases, the ECG signal is less prone to deformation than other biometric modalities in a hospital environment. In cardiovascular disease conditions also, the deformation occurs for a limited period of time. Consequently, ECG-based biometrics present a practical choice not only for healthcare services but also for an array of diverse applications, owing to its ease of acquisition [10].

This thesis work is aimed towards developing a novel framework for person-adaptive cardiovascular disease diagnosis system. Figure 1.1 (a) represents the conventional global CVD diagnosis system, while Figure 1.1 (b) outlines the fundamental block diagram and constituent modules of the proposed personalized CVD diagnosis framework. To this end, we have designed efficient ECG based biometric systems for extracting robust person specific information and biometric application. Secondly, we have designed person independent global CVD diagnosis systems for better CVD diagnosis by exploiting the spatio-temporal variation in the multi-lead ECG signal. Finally, we introduced a person-adaptive diagnostic framework, infusing the person-specific biometric information into the diagnostic network.

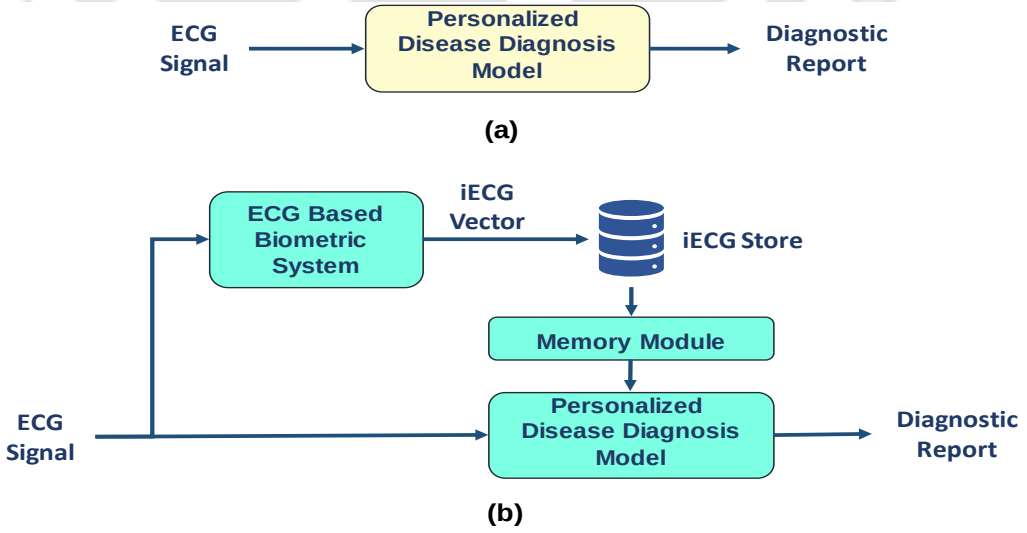

Figure 1.1: (a) Global person independent CVD diagnosis system. (b) Proposed person-adaptive CVD diagnosis system.

The remainder of this chapter is structured as follows: We begin by providing an introduction to the ECG signal, offering insights into its morphological characteristics and the electro-physiological

processes of the heart (Section 1.1). Next, we delve into a concise exploration of both multi-lead and single-lead ECG signals, as well as the pathological manifestations of cardiac diseases (Section 1.2). Section 1.5 provides a comprehensive review of existing ECG-based biometric systems. In Section 1.6, we present a brief survey of existing global and personalized CVD diagnosis systems. Our work's motivation and contributions are elaborated upon in Section 1.7 and Section 1.8. Finally, the organizational structure of the thesis is outlined in Section 1.9.

1.1 ECG Morphology: An Insight into Cardiac Activity

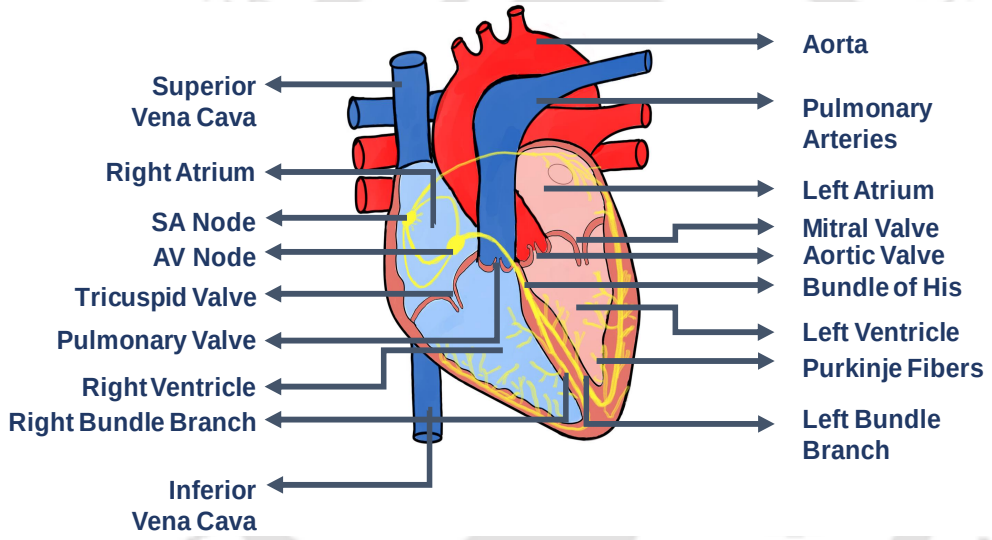

Figure 1.2: Cross-Sectional Anatomy of the Human Heart

The ECG signal is a graphical representation of the heart's electrical activity. It records the propagation of electrical impulses throughout the cardiac muscle, which in turn coordinates the highly synchronized contraction of cardiac muscle fibers. This results in heart's effective pumping action enabling the circulation of blood for oxygenation in the lungs and distribution throughout the body. As depicted in Figure 1.2, the heart consists of four chambers (left atrium (LA), right atrium (RA), left ventricle (LV), and right ventricle (RV) and four valves (tricuspid, mitral, semilunar, and aortic) that regulates the blood flow. The cardiac stimulus generates rhythmically by the pacemaker cells in sino-atrial (SA) node and spreads across the LA and RA leading to atrial contraction. This phenomena is known as atrial deploraisation that generates the P-wave. Subsequently, a part of the cardiac pulse reach the atrioventricular (AV) junction which works as a relay connecting the atria and ventricles. From here, the electric stimulus spreads through the left and right bundle branches. This

stimulates the intraventricular septum followed by the ventricular contraction. This event is marked by a large QRS complex. The depolarization of ventricles pumps the deoxygenated blood to lungs and oxygenated blood all across the body, followed by ventricular repolarization. The ventricular repolarization is marked by the T-wave followed by the ventricular diastole phase. Figure 1.3 depicts the cardiac cycles and phases. Below, we have described some of the important morphological features.

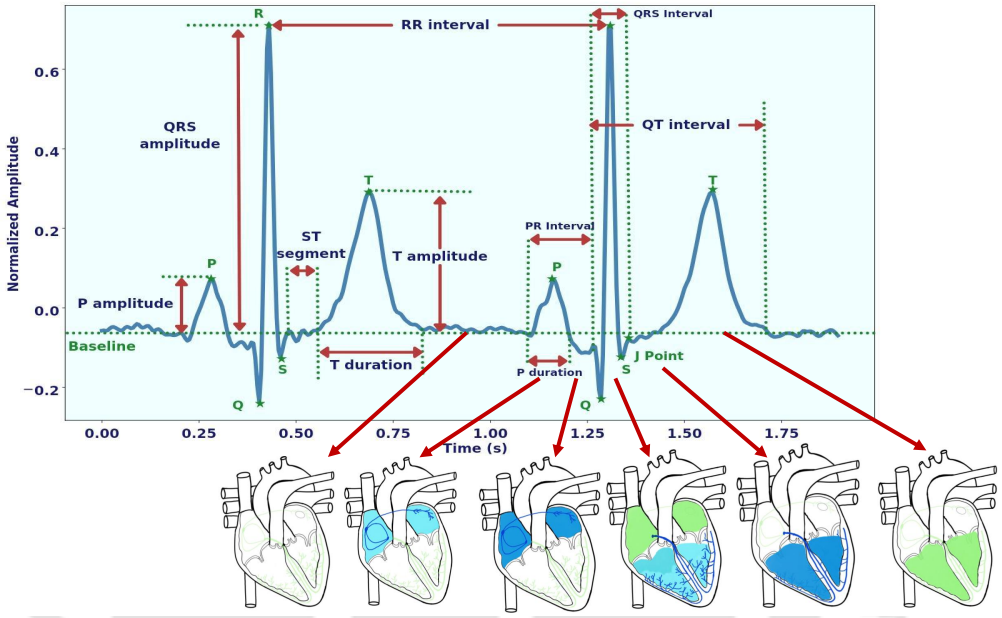

Figure 1.3: Basic ECG morphology and its relationship with heart's depolarization and repolarization sequence

P Wave: The P wave represents the cardiac impulse responsible for atrial depolarization. Typically, the duration of P wave is approximately 80 milliseconds, and its shape is generally upright, except in lead aVR. Cardiac morbidity affecting atria may result in deformed P wave.

QRS Complex: The QRS complex is the consequence of the spread of the cardiac pulse across ventricles and atrial repolarization. The QRS Complex comprises of Q wave, R wave, and the S wave. The initial negative deflection, i.e., Q wave is the result of spread of cardiac impulse across the intraventricular septum. The R wave, a positive spike, signifies the strong cardiac impulse responsible for stimulating the thick ventricular muscles, followed by a negative S wave. The average duration of a QRS complex is approximately 100 ms.

ST Segment: A normal ST segment appears as an isoelectric line, occasionally exhibiting slight elevation or depression in certain instances. The ST segment corresponds to the duration between the cessation of ventricular depolarization and the onset of ventricular repolarization. The junction between the beginning of ST segment and the end of QRS complex is referred to as J point. The ST

segment is a major daignostic cue for several cardiac diseases, such as myocardial infarction (MI), bundle branch block (BBB).

T Wave: The asymmetrical T wave represents the ventricular repolarisation. On average, the duration of T wave is approximately 160 ms. Generally, the T wave exhibits an upright orientation, with exceptions noted in leads V1 and aVR.

RR Interval: The RR interval represents the time duration between two consecutive heart beats. The inverse of the RR interval, namely heart rate is an important diagnostic parameter for cardiac health monitoring. A healthy cardiac rhythm within 60 beats per minute to 100 beats per minute is called sinus rhythm. A sinus rhythm below 60 beats per minute is sinus bradycardia and above 100 beats per minute is sinus tachycardia.

1.2 ECG Leads

The pumping action of the heart is regulated by a rhythmic electric impulse generated at the SA node and its spread throughout the heart. The ECG signal is a graphical recording of this electric impulse's propagation through the body over time, captured using electrodes. Since human body acts as a conductor of electricity, the ECG signal is recorded by placing electrodes at various locations of the body surface, such as chest, ankles, or wrists. The ECG signal recorded at an instant represents the sum total of uncanceled potential of heart cells. An ECG lead is a vector representation of a single equivalent dipole of heart's electromotive force recorded using two electrodes. There exists two types of ECG leads: bipolar and unipolar. In the case of bipolar leads, the elctrodes face the sites with similar potential difference, while in unipolar leads the potential variation of one of the electrode is negligible compared to the other one. The morphological characteristics of the ECG signal recorded from different parts of the body depends on the cardiac impulse generator (i.e. action potential generated by the SA), the characteristics of volume conductor, and the spread of excitation.

Several methods for recording ECG signals have evolved since the early 20th century. These methods employ various acquisition protocols, advocating distinct electrode placement configurations to capture the heart's electrical impulses' spatial progression, thus enhancing diagnostic precision [11–15]. Among these approaches, the gold standard 12-lead ECG recording setup and the holter ECG recording find wide utility in clinical diagnostics. Holter ECG recordings, in particular, serve the purpose of continuous, long-term cardiac health monitoring. Recent advances in wearable

technologies, such as wristwatches and chest straps, have facilitated the acquisition of single-lead ECG data from the upper body limbs or the chest. These single-lead ECG recordings are instrumental for continuous monitoring of cardiac health, particularly for the detection of arrhythmic disorders, as well as for biometric applications. The ensuing sections delve into the diversity of ECG acquisition setups and the characteristics of the recorded ECG signals.

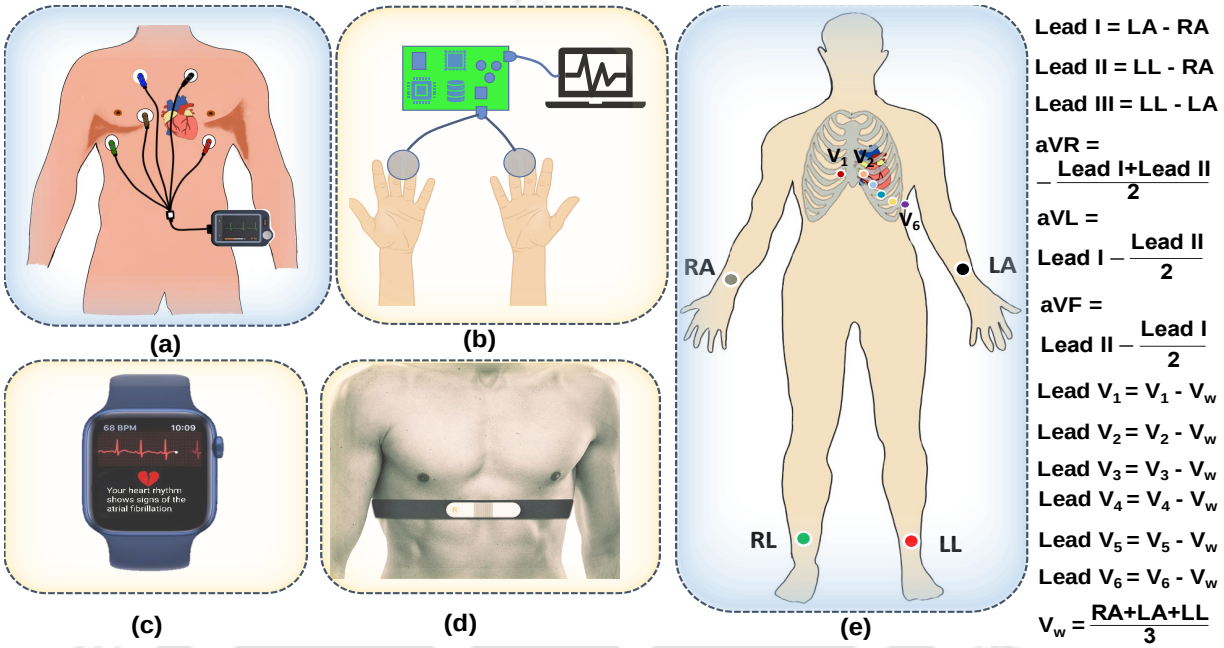

Figure 1.4: A schematic diagram depicting various ECG recording setups. (a) Showcases the Holter ECG recording system, (b) Demonstrates the *off-the-person* ECG recording setup, primarily used for biometric applications, (c) Smart watch for recording ECG signal and other cardiac vitals, (d) Features a chest band designed for recording ECG signals, (e) Illustrates the gold standard 12-lead ECG recording setup.

1.2.1 Single-Lead ECG

Single-lead ECG recording methods offer practical advantages for long-term health monitoring. They can be acquired with ease, involving only two electrodes, and demand minimal storage space. While single-lead ECG signals may not suffice for in-depth cardiac diagnostics, they provide valuable insights into long-term cardiac health, arrhythmic disorders, emotional and physical parameters, as well as biometric information. For reference, please see Figure 1.4, which presents a schematic diagram of various ECG recording procedures. Among these, the Holter ECG recording process, depicted in Figure 1.4 (a), is commonly recommended by clinicians for the extended recording of single or multi-lead ECG signals. Prolonged ECG recordings are pivotal for the early detection of occasional diagnostic signatures. Recent advancements have yielded compact wearable devices,

such as wristwatches and chest straps, illustrated in Figure 1.4 (c) and (d).

Single-lead ECG signals also play a pivotal role in biometric applications, offering comfort and convenience that enhance user acceptability. Researchers have developed various acquisition setups that reduce the reliance on conducting gels, facilitating the easy acquisition of ECG signals [10, 16–19]. These acquisition setups are referred to as *off-the-person* settings, as depicted in Figure 1.4 (b). Most of these single lead recording setups are designed to capture the conventional Lead I ECG signal. In the following section, we will provide a comprehensive overview of the standard 12-lead ECG recording setup used in clinical settings.

1.2.2 Multi-Lead ECG

Since the early 20^{th} century, multiple multi-lead ECG recording configurations have emerged for cardiac disease diagnosis. These methods differ in the number of electrodes used and their placement on the body's surface. The spatial resolution of the heart's electrical activity improves with an increase in the number of electrodes. The first practical ECG recording setup was developed by Einthoven, where he used three bipolar limb leads; *Lead I, Lead II*, and *Lead III* [11]. He conceptualized these limb leads as forming an equilateral triangle, with the heart at its center. Einthoven's law equated the magnitude of the potential in *Lead II* to the sum of potentials in *Leads I* and *III*. Wilson subsequently introduced unipolar leads employing a virtual central terminal, often referred to as Wilson's central terminal (V_w) [20]. V_w is calculated by connecting the limb leads. The unipolar leads are derived by measuring the potential difference between any location over the body and V_w. Goldberger later introduced the unipolar augmented limb leads: , aVR, aVF, and aVL [14]. When combined with the bipolar limb leads, this configuration views the heart's electrical activity in the frontal plane. Precordial chest leads, commonly known as V_x (where $x \in 1..6$), were introduced to view the heart's electrical activity in the horizontal plane [13]. The combination of the six limb leads and six precordial chest leads forms the gold standard 12-lead recording system, illustrated in Figure 1.4(e).

With the advent of volume conductor theory, Earnest Frank introduced the vectorcardiogram (VCG) or Frank XYZ system [12]. This system employed eight electrodes for ECG recording. Subsequently, Mason and Likar developed the Mason-Likar (M-L) lead system for easily recording the twelve-lead ECG signal while exercising [15]. The choice of a particular lead system is guided by the desired clinical information, clinical issues, and practical considerations.

The 12-lead ECG signals provide a comprehensive three-dimensional perspective of the heart's

electrical activity. In Figure 1.5(a), you can see a detailed three-dimensional representation of the 12-lead ECG system. As illustrated in Figure 1.5(a), lead I, lead II, and lead III provide views of the heart's electrical activity in the frontal plane at orientation angles of $0°$, $60°$, and $,120°$, respectively. Similarly, lead aVL, aVR, and lead aVF are oriented at $-30°$, $-150°$, and $90°$, respectively. The chest leads capture the heart's electrical activity from a horizontal or cross-sectional perspective. The left lateral portion of the heart is observed through lead aVL, lead I, lead V5, and lead V6, while the right lateral view is obtained using lead aVR. The inferior aspect of the heart's electrical activity is presented through Lead II, lead aVF, and Lead III. The anterior portion of the heart's electrical activity is captured from various angles through lead V1, lead V2, lead V3, and lead V4. Depending on the precise CVD subtype and the anatomical region affected, distinct pathological markers are evident in specific ECG leads. The variation in the morphological shape of an ECG waveform across different leads carry substantial clinical information [21].

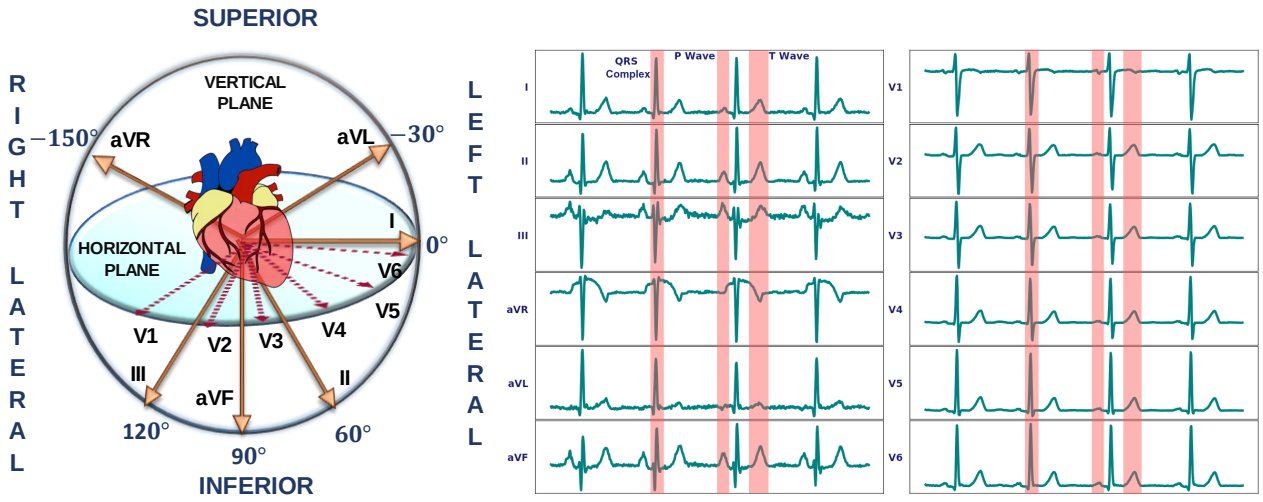

Figure 1.5: (a) Three dimensional view of heart's electrical activity as observed by 12-lead ECG recording system. (b) 12-lead ECG recording of a healthy person

1.3 Sinus ECG Signal

The sinus ECG signal represents the electrical activity of a healthy heart, originating from the SA node, the heart's intrinsic pacemaker. Figure 1.5(b) presents a 12-lead ECG recording taken from a

healthy subject. Diagnosing cardiac conditions using ECG signals relies heavily on the interpretation of distinctive features within specific waveforms, particularly the P wave, QRS complex, and T wave. An overview of the morphological characteristics of these waveforms within a sinus ECG signal can be found in Table 1.1 and Table 1.2.

Atrial depolarization, characterized by the P wave, commences with the spontaneous depolarization of pacemaker cells within the sinus node. The typical trajectory of atrial depolarization directs electrical activity downward towards the left leg, approximately at an angle of $+60^{\circ}$ relative to the frontal plane. Consequently, lead aVR consistently records a negative P wave, while lead II registers a positive wave.

Subsequently, atrial depolarization is followed by ventricular depolarization, leading to the formation of the QRS complex. Ventricular depolarization comprises two principal phases: the initiation of interventricular septum stimulation and the simultaneous depolarization of both ventricles. The phase associated with septal stimulation is relatively brief, lasting less than 0.04 seconds. During this phase, the net electric dipole is represented by an arrow pointing toward the right ventricle. This results in a small positive r wave in lead V1 and a negative q wave in lead V6. In the second phase of ventricular depolarization, the electrical force of the left ventricle predominates, resulting in a net dipole pointing in the direction of the left ventricle. Consequently, a negative deflection appears in the right chest leads (S wave), while a positive deflection manifests in the left chest leads (R wave). The shape of the QRS complex varies from a deep negative S wave to an R wave as we progress from lead V1 to V6 across the chest. This increase in the height of the R wave, typically peaking around lead V4 or V5, is termed normal **R wave progression**. The point at which the amplitude of the R wave equals that of the S wave is referred to as the **transition zone**. Since the electrical force is directed toward the left ventricle, lead aVR predominantly displays a negative QRS complex. However, the morphological shape of a healthy QRS complex may vary among individuals, depending on the orientation of the mean QRS axis. Some individuals exhibit a qR complex in leads III, II, and aVF, along with an RS complex in leads aVL and I. Conversely, others may display an rS complex in leads III and aVF, and a qR complex in leads I and aVL.

As previously discussed in Section 1.1, the ST segment represents the early phase of ventricular repolarization. In its typical form, the ST segment maintains an isoelectric profile. Deviations in the ST segment hold crucial clinical significance for diagnosis. However, few individuals with normal cardiac

function may exhibit ST segment elevation due to early ventricular repolarization. Subsequently, the ventricular depolarization phase is characterised by the T wave. The deflection in the T wave typically aligns with the main QRS deflection. The T wave is consistently positive in lead II and leads V4 to V6, while it appears negative in lead aVR. In the right chest leads, the normal T wave may exhibit negativity, isoelectricity, or positivity. Furthermore, if the T wave is positive in any chest lead, it must consistently maintain positivity in all chest leads situated to the left of that specific lead; otherwise, it is considered abnormal [21].

Table 1.1: Morphological characteristics of normal P wave, QRS complex, and T wave in limb leads

Waveform \ Lead	Lead I	Lead II	Lead III	Lead aVR	Lead aVL	Lead aVF
P wave	upright	upright	negative/upright	negative	negative	upright
QRS complex	upright	upright	negative/upright	negative	upright/negative	negative/upright
T wave	upright/negative	upright/negative	upright	upright	upright	upright

Table 1.2: Morphological characteristics of normal P wave, QRS complex, and T wave in chest leads

Waveform \ Lead	Lead V1	Lead V2	Lead V3	Lead V4	Lead V5	Lead V6
P wave	upright/biphasic	upright/biphasic	upright	upright	upright	upright
QRS complex	negative	equiphasic/negative	equiphasic/upright	equiphasic/upright	upright	upright
T wave	negative/upright	upright	upright	upright	upright	upright

1.4 Cardiovascular Diseases: Pathological Manifestation in ECG

In this section, we have provided a concise overview of different cardiovascular diseases (CVDs) and their corresponding pathological manifestations within the ECG signal. The discussion encompasses major CVDs, including myocardial infarction (MI), bundle branch block (BBB), hypertrophy (HYP), atrial fibrillation (AF), supraventricular arrhythmia (SVARR), and ventricular arrhythmias.

1.4.1 Myocardial Infarction

The myocardial cells require oxygen and other nutrients for heart pumping and other functions. Myocardial Infarction (MI) is a condition that results due to the occlusion of one of the coronary arteries supplying the heart with the oxygenated blood and nutrients. The oxygenated blood is supplied through three main coronary arteries and their branches. The right coronary artery (RCA) supplies

the posterior and a part of the lateral wall of the heart. The left main coronary artery (LCA), which is short, splits into branches into left anterior descending (LAD) and left circumflex (LCx) coronary arteries. The LAD artery supplies the ventricular septum and the anterior wall of left ventricle, where as the LCx artery supplies to the lateral wall of the left ventricle. The progressive atherosclerotic plaques and subsequent thrombyotic deposition in the arteries may lead to the partial or complete occlusion of one or more coronary arteries. This, in turn, gives rise to MI, a process characterized by three progressive phases: myocardial ischemia, acute ischemia, and necrosis (myocardial infarction). The pathological characteristics that manifests in different leads depends on the location of the necrosis and occlusion within the coronary arteries. Consequently, the MIs can be classified into inferior MI (IMI), infero-lateral MI (ILMI), anterior MI (AMI), antero-lateral MI (ALMI), and antero-septal MI (ASMI). The pathological characteristics also vary depending on the severity of the ischemia, i.e. subendocardial ischemia or transmural ischemia.

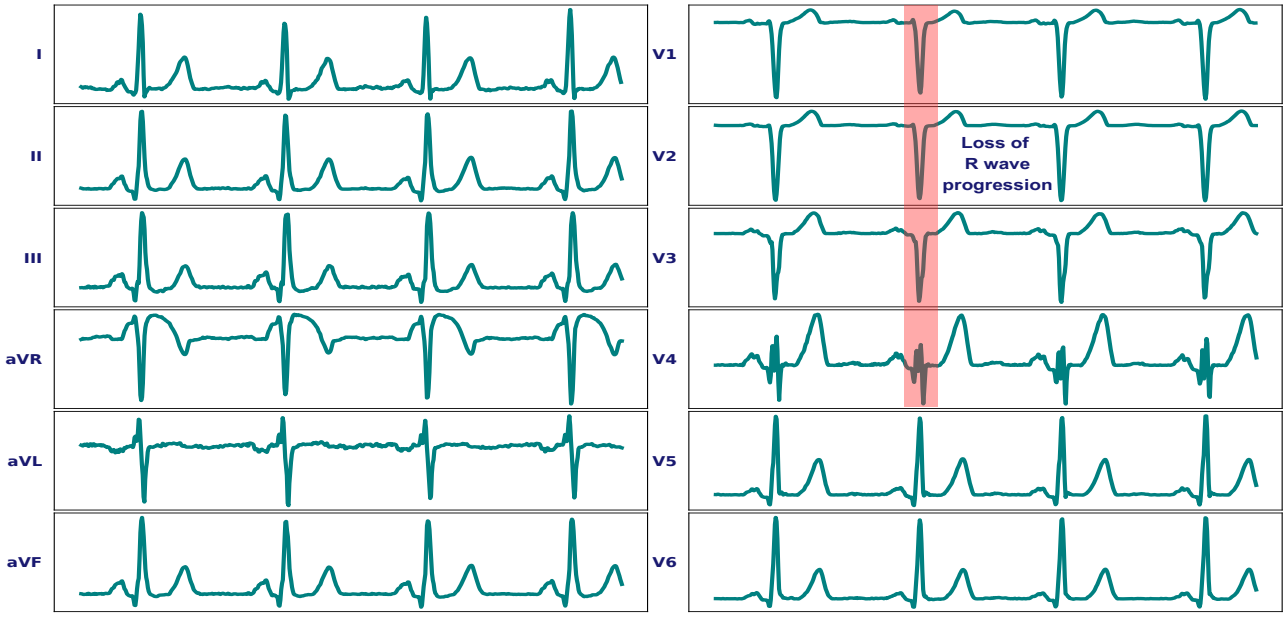

Figure 1.6: 12-lead ECG of a subject suffering from ASMI.

Anterior MI: Anterior MI is primarily caused by the occlusion of the LAD coronary artery, responsible for supplying the heart's anterior free wall. On an ECG, the signature of AMI typically includes ST-segment elevation or loss of normal R wave progression in precordial leads, notably in V1 to V6. Furthermore, the presence of abnormal Q waves, as part of QS or QR complexes, are generally observed in leads V3 to V6 due to the AMI.

Antero-Septal MI: The Antero-Septal MI results from damage to the tissues located around antero-

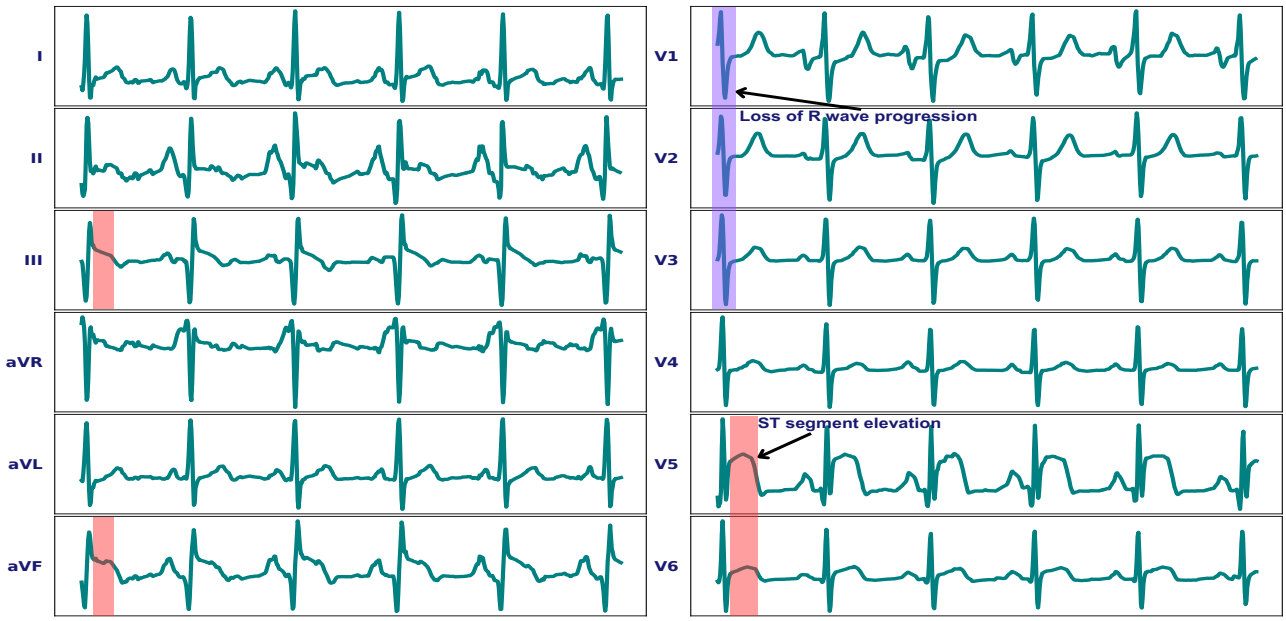

Figure 1.7: 12-lead ECG of a subject suffering from ALMI.

septal wall (region between the left and the right ventricles). This damage leads to a loss of septal depolarization voltage, causing the disappearance of R waves in leads V1 to V3 and the emergence of entirely negative QS complexes. Additionally, ST-segment elevation may be observed in various ECG leads. These pathological characteristics can be observed in the 12-lead ECG signal of a patient with ASMI, as presented in Figure 1.6.

Antero-Lateral MI: Antero-lateral MI results from the combined occlusion of the left anterior descending artery and the right coronary artery. In a multi-lead ECG, ALMI is characterized by pathological indicators,i.e., ST-segment elevations and T-wave inversion in lead I, lead aVL, lead V3, lead V4, lead V5 and lead V6. Additionally, a loss of r-wave progression may be seen in lead V2.

Inferior MI: The Inferior MI is generally caused due to the occlusion of RCA or less commonly due to LCx coronary artery occlusion. Pathological charecteristics can be, such as the abnormal Q-waves, ST segment elevations appear in lead II, lead III and lead aVF.

Infero-lateral MI: The infero-lateral MI is due to the combined occlusions in the right marginal artery and the LAD coronary artery. The pathological signatures of the ILMI includes the presence of abnormal Q-wave and the ST-segment elevations in lead III, aVL, aVF, V5 and V6. Figure 1.7 shows the ECG signal of a patient with ILMI.

1.4.2 Bundle Branch Block

Bundle branch blocks (BBB) are the ventricular conduction disturbances resulting from the blockage in the electrical pathways, i.e., right or left bundle branch. In the case of a right BBB (RBBB), a delay in right ventricular stimulation ensues. This delay results in a pathological rSR' complex characterized by a broad R' wave in lead V1, and lead V6 demonstrates a qRS-type complex with a broad S wave. It's important to note that RBBB can occur as an isolated ECG finding, often without any underlying heart disorder, thus not inherently indicative of heart disease itself. The left BBB (LBBB) presents distinct characteristics, prominently showcased by a widened QRS complex, different from RBBB in morphological characteristics. In LBBB, the entire ventricular stimulation process primarily directs its forces towards the left chest leads. Notably, in lead V1, a small notching may appear at the nadir of the QS wave, and lead V6 may exhibit notching at the peak of the broad R wave. In summary, diagnosing a complete LBBB pattern is highly probable when a wide QRS complex, usually exceeding 120 milliseconds, is observed. Additionally, LBBB can lead to T-wave inversion in the left precordial leads. Figure 1.8 illustrates a 12-lead ECG signal of complete LBBB, providing a visual representation of the discussed pathological characteristics in the ECG signal. Complete LBBB is often a diagnostic indicator for several previously undiagnosed clinically significant structural abnormalities. These can include advanced coronary artery disease (CAD), valvular heart disease (such as mitral and/or aortic conditions), hypertensive heart disease, and underlying cardiomyopathy. Identifying RBBB can serve as an initial clue to these potentially serious health concerns.

1.4.3 Cardiac Enlargement

Cardiac enlargement, or the expansion of one or more heart chambers, occurs due to an increase in cavity volume, wall thickness, or a combination of both factors. This enlargement typically arises from chronic pressure or volume loads on the heart muscle. Pressure loads, as seen in cases of systemic hypertension or aortic stenosis, lead to an increase in wall thickness, resulting in wider muscle cells—a condition known as concentric hypertrophy. On the other hand, volume loads, which may be due to factors such as valve regurgitation or dilated cardiomyopathy, primarily cause ventricular and atrial dilation. In cases of dilation, the heart muscle cells tend to elongate, a phenomenon referred to as eccentric hypertrophy. The cardiac enlargements can be broadly classified into four categories: right atrial overload (RAO), left atrial overload (LAO), right ventricular hypertrophy (RVH),

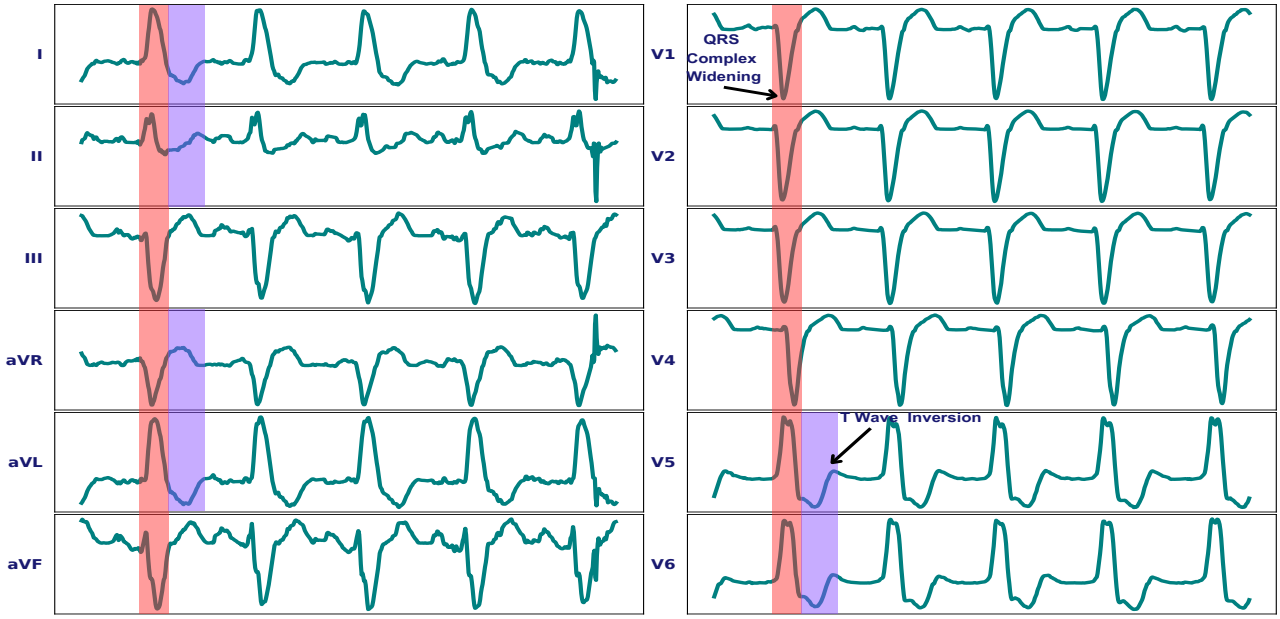

Figure 1.8: 12-lead ECG of a subject suffering from CLBBB.

and left ventricular hypertrophy (LVH). Pathologic hypertrophy generally result in scarring and change in myocardial geometry, which may further lead to various arrhythmia and chronic heart failure (CHF).

Right Atrial Overload (RAO): RAO typically arises as a result of right atrial overload, often associated with tricuspid valve and pulmonary valve diseases. This condition manifests in the ECG with a high amplitude P-wave exceeding 0.25 mV in leads II, III, and aVF, and a deep P-wave in lead V1. The duration of the P-wave generally remains unaffected in RAO.

Left Atrial Overload (LAO): LAO is frequently associated with left atrial overload, often related to mitral valve disease. Recognizable ECG signs include notched P-waves and biphasic P-waves with negative amplitudes exceeding 0.1 mV, typically observed in leads II and V1, respectively. LAO often leads to a prolongation of atrial depolarization, resulting in a broader P wave.

Right Ventricular Hypertrophy (RVH): RVH typically develops as a result of pulmonary valve stenosis and pulmonary hypertension. ECG indicators of RVH include tall R-waves with amplitudes exceeding 0.7 mV in leads V1 and V2. Additionally, wide R-waves are observed in leads V5 and V6. RVH affects both depolarisation as well as repolarisation. Thus, T wave inversion may be seen in right precordial leads.

Left Ventricular Hypertrophy (LVH): LVH is frequently associated with aortic valve and mitral valve diseases. Detecting LVH is clinically important, since it is a sign of potentially life threatening overload

state and shows a potential CHF. Pathological ECG features include tall R-waves in leads I, V5, and V6, as well as deep S-waves in leads III, V1, and V2. These ECG findings signify LVH. ST-T changes are also often seen in the case of LVH. Figure 1.9 shows a 12-lead ECG signal of a subject with LVH.

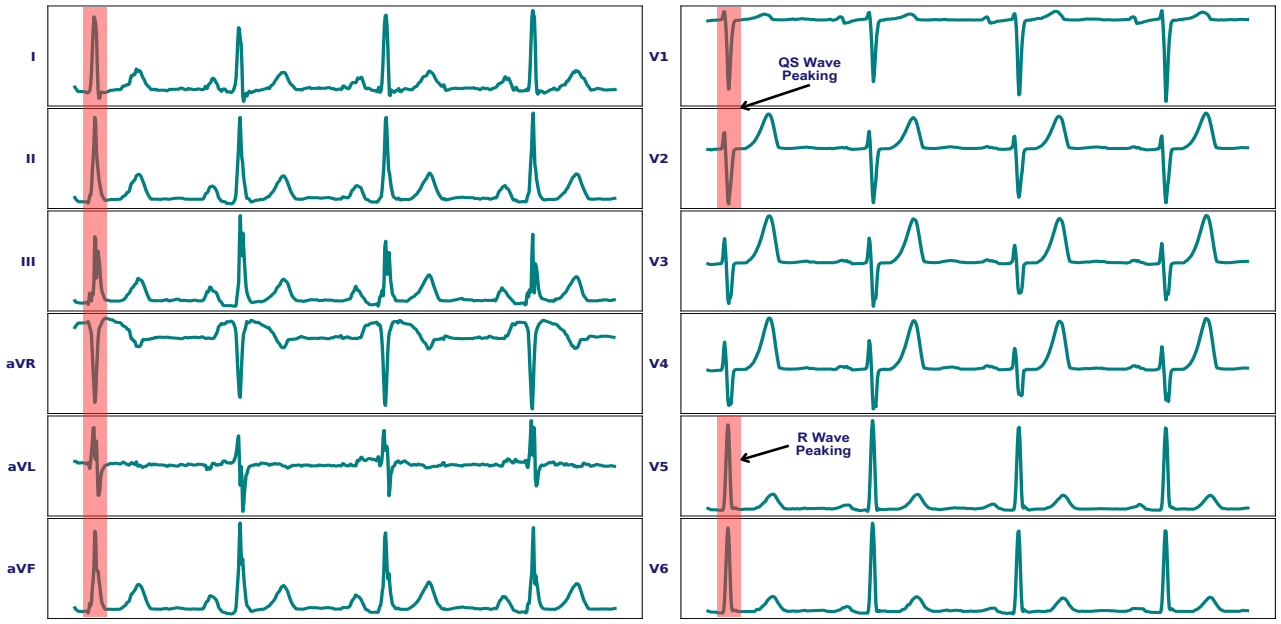

Figure 1.9: 12-lead ECG of a subject suffering from LVH.

1.4.4 Supraventricular Arrhythmia

Supraventricular arrhythmias constitute a group of heart rhythm disorders initiated by ectopic beats originating at the AV junction (premature AV junctional complexes (PJCs)) or atria (premature atrial complexes (PACs)). These ectopic beats trigger arrhythmias through either focal or reentrant mechanisms. The focal mechanism involves repetitive firing of a non-sinus pacemaker, while the reentrant mechanism results from the non-uniform spread of a depolarization wave due to unidirectional blockage in one of the pathways. Below, we offer a brief overview of some significant supraventricular arrhythmias.

Atrial Fibrillation (AFIB): Atrial fibrillation is characterized by chaotic atrial electrical activity that lacks a consistent and stable circuit in the atria. AFIB typically originates in the vicinity of the pulmonary vein–left atrial junctions, involving the emergence of rapidly firing ectopic foci and progressive involvement of more atrial tissues, forming multiple unstable micro-reentrant circuits. Atrial electrical activity in AFIB manifests as irregular, continuous, and disorganized fibrillatory (f) wavelets in ECG. These wavelets exhibit variations in amplitude, polarity, and frequency. Notably, lead V1 is often

the most informative lead for identifying the irregular atrial activity characteristic of AF. Irregular f waves are most pronounced in this lead. AFIB increases the risk of stroke due to potential blood clot formation in the fibrillating atria and may lead to heart failure if left untreated.

Atrial Flutter (AFLT): Atrial flutter is a major tachycardia, which is associated with increased risk of thromboembolism, warranting specific therapeutic considerations. This arrhythmia typically involves a macroreentrant circuit located in the right atrium, with impulses traversing through the atrium from top to bottom. Cardiologists further categorize AFLT as typical, where the circuit involves the cavo-tricuspid isthmus, and atypical, where scar tissue in the atria is often the substrate for the circuit. Atypical atrial flutter is less common and may involve scar tissue in the left or right atrium.

Paroxysmal Supraventricular Tachycardia (PSVT): PSVT refers to intermittent episodes of rapid heartbeats that originate above the ventricles. It includes various distinct arrhythmias, such as atrioventricular nodal reentrant tachycardia (AVNRT), atrioventricular reentrant tachycardia (AVRT), mono-focal, and multi-focal atrial tachycardia. AVRT involves an extra electrical pathway between the atria and ventricles, while AVNRT results from abnormal electrical circuits within the heart's AV node. In cases of focal tachycardia, an ectopic atrial focus autonomously generates rapid electrical impulses. These episodes may commence and cease abruptly and can be triggered by diverse factors. It's important to note that in patients with underlying coronary artery disease, PSVT may potentially induce myocardial ischemia, leading to angina or, in susceptible individuals, congestive heart failure (CHF).

1.4.5 Ventricular Arrhythmia

Ventricular arrhythmias encompass a spectrum of abnormal heart rhythms that originate in the ventricles or adjacent structures, such as valvular outflow tracts or the fascicular system. These arrhythmias are often triggered by premature ventricular complexes (PVCs) that can arise in either the right or left ventricle. PVCs result in asynchronous and aberrant stimulation, leading to a wide QRS complex with a duration exceeding 0.14 seconds. They may occur in different combinations, including couplets (two in a row) and ventricular tachycardia (VTAC, three or more in a row), during which the heart rate varies between 100 and 250 beats per minute. If a VTAC episode persists for more than 30 seconds, it may progress to ventricular fibrillation (VFIB), the most severe form, potentially causing cardiac arrest. In VFIB, the ventricles lose their ability to effectively pump blood into the lungs and arteries due to unsynchronized contractions.

In certain cases, isolated PVCs occur so frequently that they appear after every normal beat, resulting in ventricular bigeminy. Repeating sequences of two normal beats followed by a PVC are known as ventricular trigeminy. Another ventricular arrhythmia is the ventricular escape rhythm (VER), which typically occurs when the SA node fails to maintain an adequate heart rate. This protective mechanism ensures that the heart continues to beat, albeit at a slower rate by shifting the heart's pacemaker activity to a subsidiary site, often located within the ventricles. VER is characterized by broader QRS complexes and delayed heart electrical activity. Ventricular arrhythmias present significant clinical challenges, necessitating timely intervention to mitigate potentially life-threatening consequences.

1.5 ECG Based Biometric Systems: A Review

In the field of biometric authentication system, the ECG signal as a unique identifier has emerged as an evolving and promising field of research. The ECG biometric system exploits the distinctive electrical patterns of the heart for identity verification and authentication applications. One of the earliest work on the ECG based biometric system was proposed by Biel *et al.* dating back to the year 2001 [22]. Subsequently, a significant body of research has been dedicated to the development of robust ECG biometric systems, geared towards practical applications. Figure 1.10 presents a schematic illustration of the ECG biometric system's overarching framework, encapsulating the five fundamental operational stages of an ECG biometric system. The biometric process starts with acquiring data from a subject followed by a series of preprocessing steps including signal denoising, segmentation and outlier rejection, and signal normalization. Subsequently, person specific representation and features are extracted using various signal processing and deep learning models. Finally, the biometric results are obtained for person identification or person verification application.

Many research works have been reported in the literature using ECG as a biometric modality. A detailed review of the existing ECG based biometric systems has been presented in [23] and [24]. Pinto *et al.* have categorised the existing works based on the ECG data acquisition process, preprocessing methods, and the feature extraction and classification approaches [23]. Here, we have presented a brief review of the existing ECG biometric systems aligning with the different operational stage as illustrated in Figure 1.10.

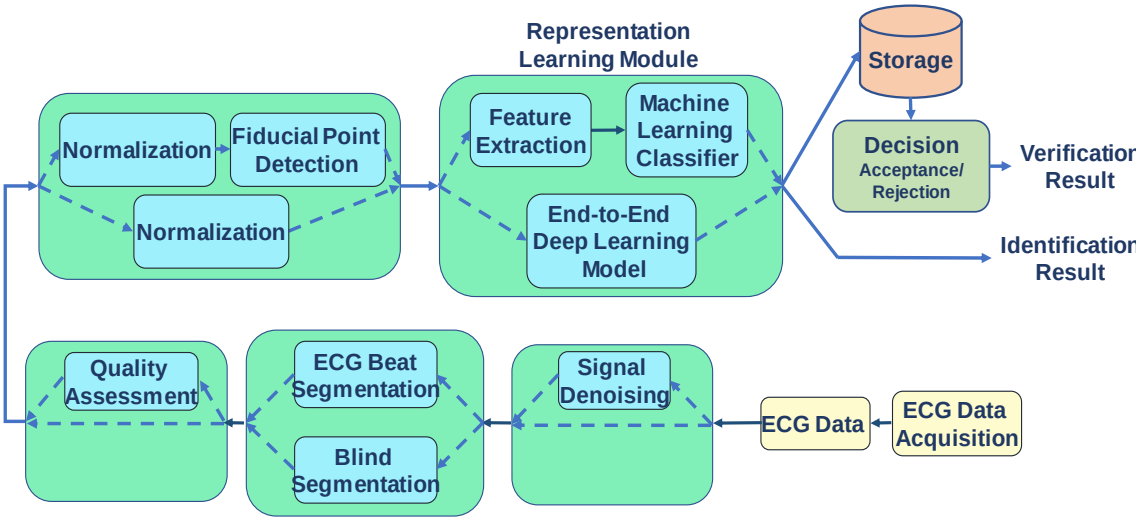

Figure 1.10: Schematic block diagram of ECG based biometric system

1.5.1 ECG Acquisition

The ECG data acquisition process for biometric applications broadly falls into two categories: ***on-the-person*** and ***off-the-person***. The *on-the-person* setup employs standard electrodes with conducting gel for signal acquisition, while the *off-the-person* setup facilitates easy signal acquisition with fewer dry electrodes and minimal skin contact.

Many of the works in literature have used *on-the-person* ECG signals in their work [23–28]. The *on-the-person* ECG signals offer a superior signal-to-noise ratio, resulting in relatively better performance. However, *on-the-person* ECG acquistion is challenging outside the hospital condition.

Off-the-person ECG signals, although characterized by a low signal-to-noise ratio, present a more practical option for biometric applications. The CYBHi database [10] and the UofTDB database [29] are the two publicly available *off-the-person* ECG database that are used in literature works. A few works have reported model performance using *off-the-person* ECG data [30–33].

1.5.2 Signal Denoising

During acquisition, the ECG signal is susceptible to various types of noise, including powerline interference, high-frequency noises, baseline wander noise, motion artifacts, and electromyographic interference, etc [34]. The nature of noise in the ECG signal varies depending on the recording procedure. ECG signals recorded in an *on-the-person* setup typically exhibit high SNR, while

those recorded in an *off-the-person* setup often contain significant high-frequency noise and abrupt changes. Many studies have employed a band-pass filter with a passband of $1-40$ Hz [30,31,35,36]. This filtering approach eliminates baseline wander noise, high-frequency noise, and powerline interference. However, it may not effectively remove high-frequency noise, particularly in *off-the-person* ECG recordings. Consequently, some studies have utilized wavelet-based methods [32,37,38], and the Savitzky-Golay filter for high-frequency noise removal [39]. Notably, a few works have sought to eliminate the filtering process entirely, as presented in [28].

1.5.3 Segmentation

Following the signal denoising step, the ECG signal undergoes segmentation to derive training or testing templates. In the existing literature, two methods have been proposed for this process, namely ECG beat segmentation and blind segmentation. The ECG beat segmentation process entails identifying the R-peak of an ECG beat and extracting a segment of the ECG signal around it. Subsequently, either a single ECG beat or a series of ECG beats from a subject is given as input to the network for modeling or validation purposes [16, 26, 30–32, 40–45]. However, this approach limits the performance of the biometric model in practical application since it depends on accurate detection of R-peak. Accurate detection of R-peak is a challenging task, particularly for *off-the-person* ECG signal. Therefore, some studies propose using a blind segmentation process to extract biometric ECG templates [28,33,46]. These methods extract a constant duration of the ECG signal without any reference point, introducing variability in the input data—a challenge for achieving improved biometric performance.

1.5.4 Normalization and Outlier Removal

The segmented ECG signal is normalized before any further operation. Two types of normalization processes are followed in general, i.e., mean and amplitude normalization, and z-score normalization. The mean and amplitude normalization is expressed in eq 1.1 and eq 1.1, respectively. The mean normalization set the mean of the signal to 0 and the amplitude normalization set the maximum amplitude of the ECG templates to 1. Here, $X \in \mathbb{R}$ represents the ECG template, μ the mean of the ECG template, and the $\bar{X}$ represents the mean normalized ECG template. The term X_{min} stands for the the maximum value in $\bar{X}$ and X_{min} stands for the minimum value.

$$\bar{X} = X - \mu \tag{1.1}$$

$$\hat{X} = \frac{\bar{X} - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{1.2}$$

The z-score normalization centers the mean of the ECG templates to 0 and scales the amplitude to achieve unit variance. It is expressed in eq 1.3. Here, the mean of the ECG template X is μ and the variance is σ.

$$\hat{X} = \frac{X - \mu}{\sigma} \tag{1.3}$$

Following the normalization process, certain studies have incorporated an outlier removal step [10, 30, 32]. Within this step, training and testing data points are excluded from the dataset based on a similarity score. This method entails calculating the mean heartbeat pattern for each subject and subsequently excluding heartbeats that deviate beyond a predetermined threshold. However, this process cannot be employed in a practical biometric system thus leading to a biased performance evaluation of the biometric model.

1.5.5 Biometric Representation Learning

The representation learning module plays a pivotal role in capturing distinct features from the ECG data of each subject for effective identification and authentication. This process involves the extraction of meaningful features, such as morphological features, rhythm characteristics, and statistical measures, from the raw ECG signals. Early works have explored various time-domain and transform domain features using signal processing techniques to extract biometric information. These features are fed to a machine learning classifier to obtain identification or verification result. However, the limitation in extracting meaningful robust features has led researchers to explore various deep learning based models to learn effective biometric representations. Recently, various convolutional neural network (CNN) and recurrent neural network (RNN) based models have been explored towards developing end-to-end identification or verification systems. The effectiveness of different representation learning techniques, their advantages, and potential limitations are briefly explored in the following subsections.

1.5.5.1 Time Domain Based Method

Early developments in ECG-based biometric systems leveraged signal processing techniques to extract meaningful biometric features from the ECG signal [23]. These early endeavors focused on the extraction of various time domain features, including amplitude, area, slope, temporal distance, and ECG waveform dynamics by detecting fiducial points [22, 32, 41, 47]. Among the earliest endeavors, Biel *et al.* employed 30 fiducial point-based features from each of the 12 ECG leads for biometric identification [22]. Huang *et al.* proposed a unified sparse representation framework that incorporates hybrid features, combining fiducial point-based features with 1DMRLBP and wavelet features [32]. Their model exhibited enhanced performance when tested with off-the-person ECG data. Additionally, Arteaga *et al.* introduced an ECG-based mobile authentication system relying on seven fiducial points identified from the ECG signal [48]. Impressively, they demonstrated improved performance with as little as 30 seconds of enrolment data and a mere 4 seconds of data for authentication. In a different vein, Lim *et al.* proposed a GMM-HMM model using time domain features as input to capture the time-varying nature of ECG morphology within the context of biometric applications [41]. Their model's performance has been evaluated using *on-the-person* ECG data. Furthermore, Huang *et al.* [49] and Cordeiro *et al.* [50] delved into the application of fiducial point-based features in real-world ECG-based biometric devices.

It's worth noting that the practical utility of these models is constrained by the limited information provided by time domain features and challenges associated with the accurate detection of fiducial points. This has led to the development of various non-fiducial point based features, eliminating the need for extensive fiducial point detection, with the exception of the R-peak [30, 44, 51, 52]. Sidek *et al.* introduced a new approach, that utilizes the convolution of ECG beats specific to each subject as the primary feature for biometric identification [44]. In [30], a 1DMRLBP feature set that can account for the temporal changes in the ECG signal, was proposed for biometric application. Their model's performance was comprehensively evaluated using both *on-the-person* and *off-the-person* ECG databases. Wang *et al.* have proposed multi-scale differential feature fused with 1DMRLBP as the base biometric features and collective matrix factorization to obtain the person specific biometric representation [51]. Another multi-scale feature based approach, which employs the autoregressive model parameters of the wavelet decomposed ECG signal as biometric features, has been proposed in [53].

1.5.5.2 Transform Domain Based Method

The time-domain based methods have primarily relied on detecting fiducial points to extract morphological characteristics, with a few different approaches that attempts to capture the time varying nature of the ECG signal. However, the challenges associated with extracting robust time-domain biometric features have motivated researchers to investigate more promising avenues within transformed domains. Transform domain based approaches employ various signal processing methods to transform the ECG signal into a different domain, where the morphological characteristics representing biometric information become more readily extractable. In this context, several studies have delved into the utilization of discrete cosine transform (DCT) coefficients derived from the autocorrelated ECG signal, coupled with linear discriminant analysis (LDA), for biometric applications [35, 54, 55]. A major advantage of these works is that they employ a blind segmentation process, mitigating the requirement of R-peak detection. In a different note, Irvine *et al.* [45] developed a principal component analysis (PCA) based method for biometric feature extraction. However, their approach require R-peak detection and ECG beat segmentation. Chan *et al.* [16] have developed a wavelet based distance measure to identify the candidate subject by computing the similarity between ECG templates. In [56], various fusion strategies have been explored to fuse biometric features from 12 lead ECG signal. In [57], Fang and Chang have proposed an unsupervised ECG based biometric identification process based on similarity and dissimilarity measure of phase space portraits. A singular value decomposition (SVD) based approach for learning biometric features along with MUSIC algorithm for improved biometric security and privacy was proposed by Wu *et al.* [42, 43]. Lee *et al.* have proposed a PCA based network using wavelet decomposed time-frequency representation of ECG signal [58].

Several studies have delved into sparse representation-based methods for the extraction of efficient biometric features from ECG signals. Wang *et al.* introduced a sparse representation technique that operates on local segments of ECG signal, achieving an impressive identification accuracy of 99.48% when tested on 100 subjects sourced from the PTB database [27]. They employed a blind segmentation process eliminating the need for R-peak detection. Some research works have explored matching pursuit approaches for biometric identification [59, 60]. Li *et al.* have presented a biometric system that combines graph regularized non-matrix factorization (GNMF) with sparse representation techniques [61]. Their method employs GNMF to enhance the discriminative properties of primary features derived from ECG beats by encoding both geometric and label information.

More recently, Tan and colleagues have investigated an innovative statistical n-best adaptive Fourier decomposition-based sparse representation framework for biometric applications [62]. In another notable study, a Hadamard code-based biometric representation was applied in the context of continual learning for online biometric systems [63].

1.5.5.3 Deep Learning Based Method

In recent years, there has been a surge in the development of deep learning based frameworks for ECG-based biometric applications. The deep learning models optimally integrate the feature extraction and classification into an end-to-end pipeline, thereby obviating the need for a separate feature extraction step. Most existing methodologies have employed CNN based models to learn effective biometric representations [26,31,33,64–67]. Silva *et al.* proposed a novel approach utilizing raw ECG beats and their spectrograms as inputs to a two-stream CNN network [31]. Their method exhibited a significant performance improvement on off-the-person ECG data, compared to machine learning based methods. In a similar vein, Zhu *et al.* introduced a low-rank fusion-based technique to learn effective biometric representations [33]. Meanwhile, Srivastava *et al.* presented a transfer learning approach, harnessing an ensemble of pre-trained deep neural network (DNN) models for biometric identification [66]. In the studies by Zhang *et al.* [64] and Labati *et al.* [26], innovative CNN-based architectures that utilize raw ECG signals as input have been introduced for biometric applications. Pinto *et al.* addressed the template cancelability and linkability problem within biometric systems by introducing a triplet loss mechanism [67].

In several other studies, researchers have performed domain transformations of the ECG signal prior to its input into CNN frameworks for learning effective biometric representation [28, 46, 68, 69]. Zhang *et al.* utilized a wavelet-decomposed multi-scale ECG signal as input to their CNN network for enhanced biometric representation learning [46]. In the study by Abdeldayem *et al.* , spectral correlation of ECG signals was employed as input to the CNN network [28]. A S-transform based method was proposed in the study by Zhao et al [69]. Furthermore, a separate set of studies has focused on generating ECG images in a two-dimensional matrix format, which are subsequently used as input to various CNN networks [70–72]. It's important to note that, except for the works in [28, 46], most of these frameworks require the precise detection of R-peaks and ECG beat segmentation.

However, the deep learning based methods lack in learning the temporal variation of the ECG signal explicitly. The temporal variation of the ECG signal preserve crucial biometric information.

This has motivated researchers to develop RNN based DL models for biometric applications. In [40], Salloum and Kuo have explored different RNN variants for ECG based biometric application. However, they have given a sequence of segmented ECG beats as an input to the RNN network. Thus, the model may not effectively learn the intra-beat variation. In [73], a bidirectional GRU based biometric system is proposed. In this work, a segmented ECG beat is given as input to the bidirectional GRU. Thus, the proposed biometric system may not effectively learn the inter-beat variation. Further, both the RNN based model and bidirectional GRU based model require R-peak detection. Allam *et al.* have proposed a combination of CNN and LSTM based framework for person authentication task [74]. In their study, they have implemented two parallel networks–one employing CNN and the other LSTM–to facilitate the learning of biometric representations. Subsequently, these networks' outputs are concatenated to derive the final authentication result. On a different note, Cherupally *et al.* have proposed an efficient hardware structure for ECG-based authentication, employing a deep neural network for representation learning [75].

1.6 Automated Cardiovascular Disease Diagnosis Systems Using ECG: A Review

Pathological changes in morphological characteristics are often manifested in the ECG signal for a majority of CVDs. Cardiologists interpret the pathological changes in ECG signals across different leads for accurate diagnosis. To aid in the diagnostic process, various automated diagnostic systems have been proposed. Figure 1.11 illustrates a schematic diagram outlining the framework of existing automated CVD diagnostic systems. This diagnostic framework comprises five core operational blocks that accept multi-lead or single-lead ECG signals as input. The initial phase involves filtering the input ECG signals to eliminate various forms of noise, sometimes followed by a quality assessment block. This quality assessment block removes the ECG signals that are heavily distorted by noise, thereby improving the accuracy of diagnostic decisions [76, 77]. Following this, ECG beats or fixed-duration ECG templates are extracted for further feature extraction and classification tasks. With the exception of the representation learning block, the preprocessing methods employed resemble those utilized in biometric applications. Below, we provide a concise overview of various existing approaches for automatic CVD diagnosis.

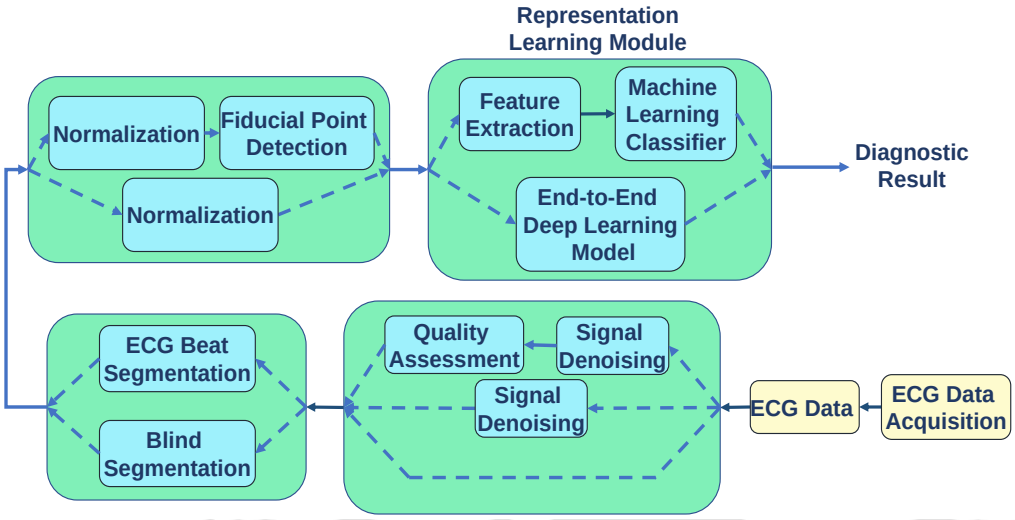

Figure 1.11: Schematic block diagram of automated CVD diagnosis systems using ECG signal

1.6.1 Diagnostic Representation Learning

Various methods have been reported in the literature for learning the diagnostic representation from both the single-lead and multi-lead ECG signal. Early works employed various signal processing technique to capture the morphological features of an ECG beat, enabling the training of machine learning models for CVD diagnosis. These works primarily focused on distinct CVD categories like arrhythmic diseases [78, 79], ischemic diseases, and myocardial infarction [80]. More recently, several deep learning-based methods have emerged for learning diagnostic representations from ECG signals [81]. While most approaches are designed for specific diagnostic categories, a limited number of studies have aimed towards diagnosing a broader range of CVDs.

1.6.1.1 Machine Learning Based Diagnostic Models

Early works have proposed different signal processing methods to capture the morphological shape of an ECG beat respective to each lead [80, 82–84]. Many works extracted time-domain morphological features to quantify the patterns that are associated with various CVDs including area and amplitudes of the R and S waves, RR interval, ST-segment deviation, QRS interval, RT segment, QT interval [85–91]. Arif *et al.* [86] used time-domain ECG features of 12 leads for classification of MI. For each beat, Q-wave amplitude, T-wave amplitude and ST-deviation parameters are extracted and combined for 12-leads. This results a 36-dimensional feature vector for each beat. Kung *et al.* have proposed an efficient light-weight arrhythmic beat detection framework for use in implantable medical

device (IMD), using random forest based classifier [88]. Dogan and Korurek proposed to use various morphological features, such as, QRS-height, QRS-width, RR-interval and kernel fuzzy c-means clustering for classification of RBBB, fusion beats, APC, and VPC [90]. In [89], Pueyo *et al.* used upward and downward slope of QRS complex as indices for the detection of myocardial ischemia.

However, evaluation of these features require accurate detection of fiducial points (P, Q, R, S and T points). The automated evaluation of P, Q, R, S, T locations along each ECG lead using signal processing algorithms are prone to error. This motivated researchers to propose various transformed domain features for effective characterisation of morphological features. Ye *et al.* have proposed a wavelet transformation along with independent component analysis (ICA) to extract morphological features for arrhythmia detection [92]. In an another approach, Raj *et al.* have proposed a discrete orthogonal stockwell transform using discrete cosine transform for efficient morphological feature extraction from time-frequency representation of ECG signal [93]. Stridh *et al.* proposed a new approach using a time-frequency distribution with a logarithmic frequency scale for characterization of atrial arrhythmia [94]. Sharma *et al.* proposed multiscale wavelet energies and eigenvalues of multiscale co-variance matrices to capture the ECG morphological features for MI detection and localisation [83]. Valverde and Arini studied the characteristic variations in the spectrum of T-wave during acute myocardial ischemia. They have obtained the T-wave spectral variance feature by using the DFT of T-wave and the spectral energy of ECG signal between 0.5 Hz to 50 Hz [95]. Liu *et al.* proposed a polynomial fitting based feature extraction technique, which fits a 20th order polynomial function defined as PolyECG-S, for detection of MI [96]. Fatimah *et al.* have proposed fourier intrinsic band functions (FIBF) based statistical features for the detection of MI from single lead ECG signal [97]. To address the challenge in detecting posterior MI from ECG signal, Khan and Pachori proposed to use derived VCG signal from 12-lead ECG signal. They employed eigen values obtained from PCA of Fourier-Bessel series expansion based empirical wavelet transform (FBSE-EWT) in their work [98]. An active learning based method is proposed to learn diagnostic representation from small amount of annotated ECG data [99]. Sun *et al.* proposed a multiple instance learning (MIL) based semi-supervised framework for MI detection [84]. Sinha and Das introduced two new features namely spectral coherence indices (SCIs) and phase coherence indices (PCIs) based on cross-spectral domain analysis for MI detection and localisation [100].

Some works have explored sparse representation based frameworks for effective pathological

feature extraction [101,102]. Abdelazez have proposed an AFIB detector using compressively sensed ECG mitigates the need for the computationally expensive process of ECG reconstruction [101]. They have extracted various statistical features by using WT, DCT, and empirical mode decomposition (EMD). Raj *et al.* have proposed a sparse representation based framework using an overcomplete gabor dictionary for effective morphological feature extraction [102].

However, most of the works in the literature lack in modeling the temporal variation of the ECG signal, explicitly. Few works have proposed hidden markov model (HMM) based models to characterise the temporal variation of the ECG signal for CVD diagnosis [87, 103–105]. Chang *et al.* have proposed a HMM based method to learn the morphological shape by exploiting the temporal variation of the ECG signal [106].

Few works have attempted to characterize the spatio-temporal variation of the ECG signal for learning effective diagnostic features [82, 107]. Fayn has evaluated spatio-temporal features from the ECG signal of lead I, II and lead V2 [82]. The decision tree classifier has been used for the detection of myocardial ischemia from the spatiotemporal features of ECG. However, their work lacks exploiting the concurrent spatio-temporal information present in the ECG signal. In [108], Padhy and Dandapat have attempted to address this shortcoming by using a higher order singular value decomposition based approach.

Most of the machine learning based frameworks available in literature are aimed towards diagnosing specific CVDs. Recently, a framework for multi-label feature selection and multi-label classification of Arrhythmia, which solves the problem of detecting multiple arrhythmia labels that might be present in a single subject [109].

1.6.1.2 Deep Learning Based Diagnostic Models

Recently, many DL based frameworks have been proposed for automated CVD diagnosis [81]. Most of the works proposed in literature have used CNN to learn the diagnostic representation. Pourbabaee *et al.* have presented a deep CNN model for the effective detection of paroxysmal atrial fibrillation detection [110]. A CNN based model for MI localisation using multi-lead ECG signals is presented in [111]. Cao *et al.* have proposed a novel lightweight neural network model using CNN for myocardial infarction (MI) detection. Their model utilises a channel-block structure to learn lead specific diagnostic information [112]. Some works have utilised the residual CNN architecture for CVD diagnosis. Han and Shi have proposed a multi-lead ResNet (MLResNet) architecture for MI detection

and localisation [113]. In [114], a ResNet based model is proposed for arrhythmia classification. Recently, some works have explored the multi-scale CNN architectures for effectively capturing the local morphological representations (i.e. shape of P, QRS, and T wave) and interval features (QRS duration, P-R interval) [115, 116]. Recently, Qin *et al.* proposed a deformable CNN to learn the multi-lead diagnostic information by adaptively changing the receptive fields [117]. However, the CNN based methods' limitation is that they do not use temporal information explicitly.

The ECG signal is a time varying signal with a repetitive ECG beat. Any alteration due to the CVDs reflects in the temporal pattern of the ECG signal. This has motivated researchers to model the temporal variation of the ECG signal [118]. Yao *et al.* have proposed a CNN-LSTM model for arrhythmia detection [119]. They have used a VGGNet architecture for learning the local representations, and an LSTM model for modeling the temporal variation of the ECG signal. In [120], Liu *et al.* have also proposed a CNN based model for learning the diagnostic representation from each ECG beat. Finally, the learned representations of each lead are aggregated using an LSTM model. Although their model utilises the inter-lead variation, it lacks exploiting the temporal variation. Hammad *et al.* have proposed a genetic algorithm (GA) based method to select the best local features learned by the DNN model. Finally, the chosen features are given to an LSTM model to learn the final diagnostic representation [121]. In [122], wavelet based features and RR intervals along with raw ECG beats are given as input to the LSTM model for arrhythmia classification. Hou *et al.* have proposed an LSTM based autoencoder model in tandem with SVM based decoder for arrhythmia classification [123]. They have shown that the LSTM model can effectively learn the diagnostic representation. In [124], an LSTM based model is proposed to learn the lead specific representation, and the final representation vector is constructed by applying self-attention across the different leads.

Recently, some works have proposed various attention based neural networks for CVD diagnosis. Gao *et al.* have proposed a novel temporal attention mechanism interleaved with CNN layers for the detection of atrial fibrillation [125]. Jin *et al.* have presented a dual attention mechanism, which is designed to learn the representations by giving attention at the intra-beat and beat level [126]. However, their model uses only single lead ECG data for diagnosing CVDs. Yoo *et al.* have proposed a xECGNet that can produce a fine-tuned attention map for multilabel classification of cardiac arrhythmia [127]. Wang *et al.* have proposed a nonlocal convolutional attention block along with a Resnet33 as the backbone neural network for arrhythmia classification [128].

1.6.1.3 Personalized Diagnostic Models

Personalized medicine (or precision medicine), has recently garnered significant research interest owing to its potential for optimizing patient outcomes through tailored treatment strategies [8, 9, 129]. In the domain of CVD diagnosis, a wide range of personalized CVD diagnostic models has been introduced with the goal of improving diagnostic accuracy [6, 130–134]. Watrous and Towell were among the pioneers in recognizing the challenge and proposing an early personalized diagnostic model [135]. Subsequently, Hu *et al.* introduced a patient-adaptive arrhythmia classification model using a mixture of expert-based approaches [130]. In their approach, they employ a two-classifier system: a global expert (GE) trained on the entire dataset and a local expert (LE) trained with a smaller, patient-specific training dataset. The final decision is determined by a gating function that dynamically weighs the classification results of both GE and LE. Similarly, Chazal *et al.* proposed a linear discriminant based two classifier model for patient-adaptive arrhythmia classfication [136]. They devised a method for combining the parameters of both the GE and LE classifier. Early works included various NN based personalized diagnostic models that exhibited significant performance enhancements [131, 137]. Jiang and Kong presented a block-based NN model utilizing an evolvable network architecture [137], while Ince *et al.* proposed an optimized NN model using multidimensional particle swarm optimization (PSO) technique [131]. During the training phase, both common and patient-specific ECG data were used to build a patient-specific diagnostic model for each subject. A similar training process was followed by Kiranyaz *et al.* to develop a 1D-CNN based arrhythmia classification model [138]. Similarly, a generalized regression NN is proposed in [139]. Llamedo *et al.* introduced a patient-adaptive arrhythmia classification model that incorporates wavelet-based morphological features and RR interval features [132]. They proposed a method that enables the GE classifier to operate independently or with varying degrees of assistance by incorporating a LE classifier based on a Gaussian mixture model (GMM). Xu *et al.* proposed a framework that feeds i-vectors, which carries the person specific information, as inputs to a DNN model alongside person-specific ECG data [140]. In contrast to previous approaches in [131, 133, 137, 138], which involve under-sampling common ECG data to improve personalized diagnostic performance, Xu and his team demonstrated that their method achieves superior performance while using the complete dataset.

The works described in the aforementioned paragraph rely on a certain amount of annotated ECG data from each subject as individual training data for adaptation. This requires expert intervention to

provide ground truth labels for every beat in the individual training dataset. However, manual expert intervention is often unfeasible due to its expense, time-consuming processes, and limited availability, rendering person-adaptive models impractical. To tackle this issue, Ye *et al.* introduced a multiview-based learning framework for automatic person adaptation using unlabeled person-specific ECG data [133]. This method utilizes the multi-view learning approach to extract a subset of high-confidence heartbeat instances corresponding to each subject for training the LE models. Similarly, Golany and Radinsky proposed a generative adversarial network (GAN) for personalized arrhythmia classification [141]. Their approach involves using the generative framework to create synthetic person-specific diseased ECG data, which are then utilized to train the person-specific diagnostic model. Zhai *et al.* presented a semi-supervised learning-based iterative framework for personalized ECG arrhythmia detection [142]. However, the majority of these frameworks train a person-specific model for each subject, leading to significant operational complexity in large-scale applications. To address this challenge, an unsupervised domain adaptation technique was proposed in Wang *et al.* (Wang *et al.* , 2021). Their approach initially trains a GE model, followed by unsupervised domain adaptation of the testing dataset using two novel objective functions, Cluster-Aligning loss and Cluster-Maintaining loss.

The works proposed in the literature require both normal and diseased ECG data of each subject for training person-adaptive diagnostic models. While a few approaches have suggested unsupervised frameworks for adaptation, they still rely on utilizing both normal and diseased ECG data specific to the subject. However, in real-world health monitoring scenarios, obtaining diseased ECG data for a subject without any prior history of cardiac disorders is impossible. Few works have been done to address this challenge. Zhou *et al.* have proposed a GAN based framework with only normal ECG signal of a subject for training a person specific model for each subject [134]. Yamaç *et al.* have proposed a zero-shot based domain adaptation framework which requires only the healthy ECG signal of a subject [143]. In a different approach, Gyawali *et al.* [7, 144] have proposed different deep generative models to disentangle the underlying generative factors leading to inter-individual variation.

1.7 Motivation

Existing literature works present a wide range of approaches for the development of robust biometric systems and CVD diagnostic methodologies,as discussed in Sections 1.6 and 1.7. However, certain research challenges remain unaddressed, which are outlined below.

- The practical application of a biometric system has certain challenges, i.e., acceptability, circumvention, and performance [145]. An easy ECG acquisition system, such as *off-the-person* recording setup, is necessary for broader acceptance of the biometric system. However, the low SNR in *off-the-person* ECG signals poses challenges in detecting fiducial points including R-peak, often used in preprocessing stage by the existing methodologies. The success of these methods depends on the correct identification of fiducial points. However, correctly identifying fiducial points is challenging, particularly for *off-the-person* ECG signals [23]. Additionally, as discussed before, the existing works lack exploiting the temporal variation of the ECG signal explicitly, for biometric application. The ECG signal, being inherently time-varying, may contain individual-specific information within the temporal variations of key ECG waveforms, including the P wave, QRS complex, ST segment, and T wave. Therefore, designing DL model explicitly focusing on learning these temporal variations within the ECG signal could substantially contribute to the acquisition of a robust biometric representation.

- The distinctive biometric traits primarily manifest in the morphological shape and the temporal dynamics of the four basic ECG waveforms, i.e. P wave, QRS complex, ST segment and T wave. The shape and duration of a waveform also vary for different ECG waveforms. Thus, the biometric information can be better modeled by learning multi-scale morphological representation. Apart from the morphology of the ECG waveforms, the temporal variation in the ECG presents crucial biometric information. The temporal variation in ECG signal is broadly of two scales, i.e. intra-beat variation and inter-beat variation [107]. Therefore, learning multi-scale local morphological shape and multi-scale temporal variation of the ECG signal is pivotal capturing effective biometric representations. Additionally, specific ECG waveform contain more biometric information corresponding to a subject. Therefore, the selective utilization of informative ECG morphology is essential to augment the biometric representation.

- In a standard clinical setup, a 12-lead ECG recording is done, where each lead views the

heart from a different angle. Thus, the 12-lead ECG signal varies across different leads as well as the time axis. This spatio-temporal variation constitutes a three dimensional view of the heart's electrical activity. The pathological characteristics due to the CVDs manifest in the form of variation in the morphological shape of the ECG waveforms across specific leads. Thus the variation of the clinical components of the ECG signal across different leads as well as along the temporal scale constitute the major cue for diagnostic decision making [21]. The existing automated diagnostic models are designed to learn the lead specific information in the first stage, and then the learned information is fused to obtain a diagnostic decision. Thus, the existing methods lack in fully exploiting the concurrent spatio-temporal variation present in the ECG signal. This has motivated us to design a neural network model that can learn the representation by leveraging the concurrent spatio-temporal variation of the ECG signal.

- Variation in the morphological characteristics of the ECG signal across different subjects leads to the degradation in the diagnostic performance of an automated CVD diagnosis system [3]. This is because the trained global models are unable to generalize to data from a new test subject, generated by a different underlying distribution compared to the training data. The existing person-adaptive diagnostic frameworks necessitate both normal and diseased ECG data for each test subject to adapt a gloabl epert (GE) model and instantiate the local expert (LE) model. In a real-world health monitoring application scenario, diseased ECG data of a subject without any past cardiac disorder won't be available. Additionally, disease data for all potential cardiac conditions may not be available for a subject, potentially limiting the diagnostic model's performance. Another practical constraint in utilizing existing methodologies is the limited storage capacity, given the necessity to store person-specific models for each subject. Furthermore, the existing person-adaptive CVD diagnostic frameworks are designed for single-lead ECG signals, whereas multi-lead ECG signals are often used in hospital environments for diagnosing cardiac diseases.

1.8 Contribution

The objective of this research is to develop an effective automated person-adaptive CVD diagnostic framework. To this end we present four major contributions towards developing an end-to-end person-adaptive CVD diagnostic framework.

- We developed a non-fiducial point based biometric system leveraging the temporal variation of the ECG signal explicitly. First we designed an LSTM based biometric system that learns the underlying temporal representation of the ECG signal. The proposed framework doesn't need the detection of any fiducial points. This is achieved by training the LSTM network with smaller segments of ECG signals as input, which are extracted by sliding a rectangular window. However, the LSTM model lacks effectively modeling the multi-scale temporal information,i.e., intra-beat and inter-beat variations present in the ECG signal. The intra-beat variation signifies the relative variations of P, QRS, and T complexes present in an ECG beat. While the inter-beat variation signifies the beat-to-beat variation. This motivated us to design a novel attention based hierarchical LSTM (HLSTM) model to learn the biometric representation by capturing the temporal variation of the ECG signal in different abstractions. This is done by stacking an LSTM layer over another LSTM layer with different update intervals. The attention mechanism of the model learns to identify ECG complexes with greater biometric relevance for each individual. These ECG complexes are given more weight to learn better biometric representation.

- A novel multi-scale temporal dynamics learning network (MSTDLNet) for capturing both the local morphological representation and multi-scale temporal dynamics of ECG signals is proposed for biometric applications. Unlike the existing works that learn the multi-scale temporal representation, we leverage the fine-to-coarse flow of information within a stacked convolutional network to learn the multi-scale temporal representation. Specifically, we have designed a convolutional kernel based multi-scale enhanced morphological representation learning (MSE-MRL) module to learn the ECG waveform's morphological representation better. Further, the multi-scale temporal dynamics of the ECG signal are learned by innovatively integrating two layers of LSTM networks at different hierarchical levels. The proposed framework offers a comprehensive approach for learning both local morphological representation and temporal dynamics of ECG signals at multiple scales.

- An attentive spatio-temporal learning based neural network (ASTLNet) to effectively learn the concurrent multi-scale spatio-temporal representation from the ECG signal is proposed. The proposed architecture consists of two modules, (i) Spatio-Temporal Representation Learning (STRL) module, and (ii) Attentive Spatio-Temporal Aggregation(ASTA) module. The STRL module consists of a clustered multi-head criss-cross attention (MHCCA) interleaved within

a hierarchical LSTM (HLSTM) network. The clustered MHCCA layer facilitates learning the spatio-temporal representation by aggregating the local temporal representations. The ASTA module is introduced to effectively aggregate the multi-scale spatio-temporal representation learned by the STRL module. The STRL module consists of two recurrent MHCCA layers followed by a novel multi-aligned attention (MAA) layer. The MAA layer is introduced to obtain multiple context vectors by giving more weight to diagnostic significant regions in the temporal dimension. The proposed model is tailored for a multilabel CVD diagnosis application, enabling the diagnosis of multiple cardiovascular diseases within a single subject.

- An unsupervised person-adaptive CVD diagnostic framework is introduced, leveraging the identity ECG (iECG) vector–a memory vector containing person-specific information. Specifically, we introduced a conditioning network for affine transformation of diagnostic model's features conditioned on patient specific information. The patient specific information are encapsulated in a knowledge space through a memory module that utilizes ECG-based biometric representation (iECG vector). For each new subject a memory vector is generated by an auxiliary attention network that leverages the person-specific information encapsulated in the knowledge space. The auxiliary attention network is trained along with the main network and the conditioning network. We have introduced an auxiliary loss to enhance the learning of a more effective person-specific representation by the memory module. The proposed framework offers a modular design, seamlessly integrated to adapt different CVD diagnosis networks.

1.9 Organization of The Thesis

This thesis work is organized into six chapters. **Chapter 1** serves as an introductory chapter, offering insights into the electro-physiological processes of the heart and the morphological characteristics of both healthy and diseased ECG signals. It also provides a concise overview of existing ECG-based biometric systems and automated CVD diagnosis methods.

The subsequent chapters (**Chapter 2** to **Chapter 5**) present the major contributions made in this work. In **Chapter 2**, a hierarchical LSTM network (HLSTM) is introduced, which is designed to learn the biometric representations by leveraging the intra-beat and inter-beat variations of the ECG signal. A comprehensive array of experiments is conducted and the ensuing results provide a thorough analysis of the impact of temporal variation on biometric representation learning.

Chapter 3 presents a novel approach for learning robust biometric features from ECG signals with low signal-to-noise ratio (SNR) by capturing the multi-scale local morphological and temporal variation of the ECG signal. Specifically, a novel neural network model is proposed by innovatively fusing CNN and LSTM networks.

Chapter 4 introduces the attentive spatio-temporal learning network (ASTLNet), designed to learn better diagnostic representation by exploiting the concurrent spatio-temporal variation of a multilead ECG signal. Extensive experimental results validate the model's effective spatio-temporal representation learning capability, leading to improvement in diagnostic performance.

In **Chapter 5**, a person-adaptive unsupervised CVD diagnosis model is proposed. The effectiveness of the proposed person-adaptive CVD diagnosis model is validated with a range of experiments. Finally a summary of this thesis work and future directions of research is provided in **Chapter 6**.

2

An ECG Based Biometric System Using Hierarchical LSTM With Attention Mechanism

Contents

Biometric systems play a vital role in ensuring the security and privacy of data-driven intelligent systems. The rapid growth of artificial intelligence (AI) based systems has spurred extensive exploration in diverse biometric modalities, such as electrocardiogram (ECG), fingerprint, face, iris, speech, gait, among others [145]. Compared to other biometric modalities, ECG signal is robust against spoof attacks due to its inherent liveliness information and less exposure to covert acquisition. The ECG signal can also be conveniently captured through *off-the-person* recording setup [10]. The ECG based biometric systems can be used in a wide range of applications, e.g., mobile devices [48], healthcare services [146, 147], and continuous authentication [30], etc. The rise in healthcare services has increased the requirement of proper security and privacy management of sensitive medical data. The ECG based biometric system can be a practical choice in the healthcare sector as it is less prone to deformation than other biometric modalities in a hospital environment.

The practical application of a biometric system has certain challenges, i.e., acceptability, circumvention, and performance [145]. An easy ECG acquisition system is necessary for the biometric system to be widely accepted. The *off-the-person* recording setup provides the simplest and the easiest way of acquiring the ECG signal. Although the *off-the-person* recording setup is the most suitable for ECG based biometric, the quality of the data is a great challenge. Thus designing a robust biometric system in the *off-the-person* recording scenario has become a necessity. Another requirement for a biometric system to be widely accepted is less time for enrolment as well as verification. However, most of the works available in the literature require 30 seconds to more than a minute of ECG data for enrolment [23, 24].

The ECG signal is a time varying signal, with biometric information embedded within its temporal dynamics. However, existing ECG-based biometric systems lack in explicitly leveraging the temporal variation for biometric applications (Section 1.3). Typically, existing works fall into two categories: fiducial point-based and non-fiducial point-based methods. Fiducial point based methods detect the R-peak or extract multiple fiducial points in the preprocessing stage. While the non-fiducial point based methods don't require the detection of any fiducial points. In fiducial point based methods, the biometric system's efficiency depends entirely on the accurate detection of fiducial points. In [30] and [31], the wrongly detected R-peaks have been identified and discarded before further processing. They have also implemented an outlier removal algorithm that removes ECG beats that are not similar to the majority of ECG beats belonging to the respective person. However, these outlier removal

algorithms cannot be implemented in a practical scenario. The non-incorporation of these algorithms will severely degrade the performance of the biometric system.

This has motivated us to develop a non-fiducial point based biometric system leveraging the temporal variation of the ECG signal explicitly. In this chapter, first we designed an LSTM based biometric system that learns the underlying temporal representation of the ECG signal. The proposed framework doesn't need the detection of any fiducial points. This is achieved by training the LSTM network with smaller segments of ECG signals as input, which are extracted by sliding a rectangular window. However, the LSTM model lacks effectively modeling the multi-scale temporal information,i.e., intra-beat and inter-beat variations present in the ECG signal. The intra-beat variation signifies the relative variations of P, QRS, and T complexes present in an ECG beat. While the inter-beat variation signifies the beat-to-beat variation. This motivated us to design a novel attention based hierarchical LSTM (HLSTM) model to learn the biometric representation by capturing the temporal variation of the ECG signal in different abstractions. This is done by stacking an LSTM layer over another LSTM layer with different update intervals. The attention mechanism of the model learns to identify ECG complexes with greater biometric relevance for each individual. These ECG complexes are given more weight to learn better biometric representation.

2.1 An LSTM Based Biometric Framework for Person Identification

This section introduces the ECG-based biometric framework proposed in this study. The block diagram of the framework is depicted in Figure 2.1. The framework consists of three core modules: (i) Preprocessing, (ii) Segmentation, and (iii) Biometric Representation Learning module. In the preprocessing module, the ECG signal is filtered from various noises, and the filtered signal is normalised. The normalised ECG signal is windowed to different segments in the windowing module. Then these segments are fed to a biometric model to learn the underlying biometric representation, followed by decision making.

2.1.1 Preprocessing

The noise present in the ECG signal varies depending on the type of recording procedure. The ECG signal recorded in an *on-the-person* recording setup has high SNR. But, the ECG data recorded in an *off-the-person* recording setup contains a significant quantity of high frequency noise and abrupt

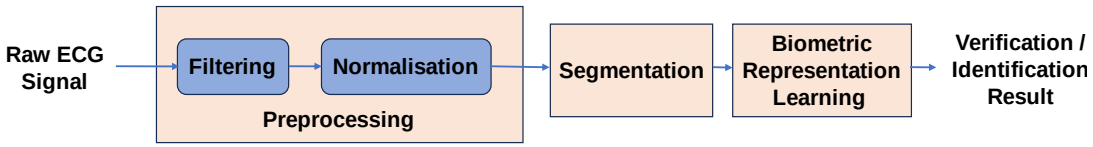

Figure 2.1: (a) Block diagram of the proposed ECG based biometric system

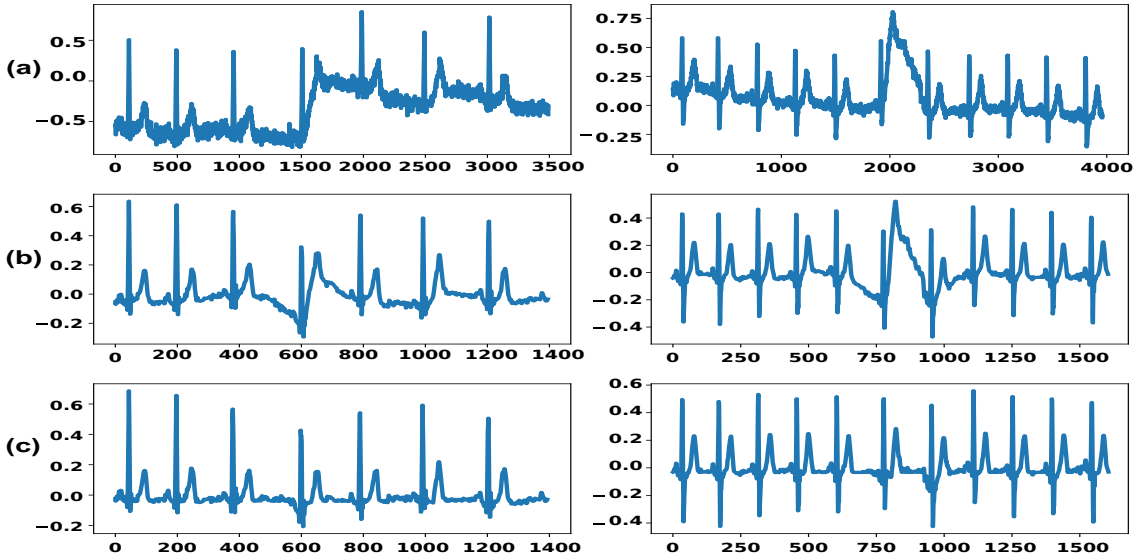

Figure 2.2: (a) Raw ECG signal of two different persons. (b) ECG signal after band-pass filtering and high frequency noise removal. It can be noticed that the noise due to sudden changes has not been filtered out. (c) ECG signal after removal of baseline drifts due to sudden changes.

changes. So, we resampled the ECG signal to a sampling frequency of 200 Hz and then passed it through a band pass filter with a lower cutoff frequency of 0.5 Hz and a higher cutoff frequency of 40 Hz [23, 31]. This filters the ECG signal from baseline wander, power line interference, and high frequency noises. Then the ECG signal is further filtered using the wavelet based method proposed by Sharma *et al.* [38]. This removes the noises that are present within 0.5 Hz to 40 Hz frequency range. Figure 2.2(a) shows two raw ECG samples, and Figure 2.2(b) shows the respective ECG signal after band pass filtering and wavelet based noise removal. From the figure, it can be observed that the sudden changes present in the ECG signal cannot be removed through low pass filtering. These noises are frequent in the case of the *off-the-person* recording setup. So, we have employed the baseline wander removal method proposed in [148] to remove the sudden changes. Figure 2.2(c) shows the respective ECG signal after the removal of sudden changes. The peak-to-peak amplitude and the mean of the filtered ECG signal are then normalised to one and zero, respectively.

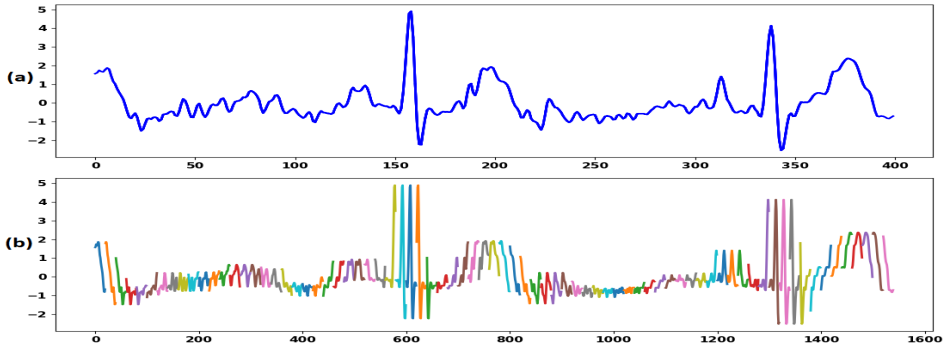

Figure 2.3: (a) One complete ECG sequence of duration 2 s. (b) The ECG segments of length 0.1 s were obtained after applying the windowing technique to the ECG sequence. The ECG segments have been plotted sequentially.

2.1.2 Segmentation

The segmentation module is the basic building block of the proposed model. In this module, the filtered ECG signal is segmented to fixed-length ECG segments by sliding a rectangular window. These ECG segments are given as input vector to LSTM cells at every time stamps. The length of the ECG segments is the same as the length of the rectangular window. In our experiments, the length of the rectangular window, i.e., T_w has been varied between 0.1 seconds to 0.8 seconds. ECG segments are obtained by shifting the window by a specified fraction of T_w. This fraction is varied from 0.25 to 0.75 in our experiments. A group of ECG segments taken sequentially forms an ECG sequence. The number of ECG segments in an ECG sequence is the same as the sequence length of the LSTM model. The number of ECG segments in an ECG sequence is chosen, such that around 2 seconds of ECG recording is needed for the same. In the windowing process first an ECG sequence is extracted from the ECG signal. The ECG sequences are extracted without any alignment with the R-peaks. This makes our model completely free of fiducial point detection. The ECG sequence is z-score standardised before segmentation. Figure 2.3 shows an ECG sequence extracted from the filtered ECG signal and the ECG segments of an ECG sequence aligned sequentially. Table 2.1 represents the number of ECG segments in one ECG sequence for different combinations of segment length and shift. To ensure sufficient training data, we have shifted every ECG sequence by a 0.25 fraction of ECG sequence length. This also works as data augmentation, which is a type of regularisation method for neural networks. While generating sequences for the test set, the ECG sequence is shifted fully without any overlap. The training and test sequences are extracted from

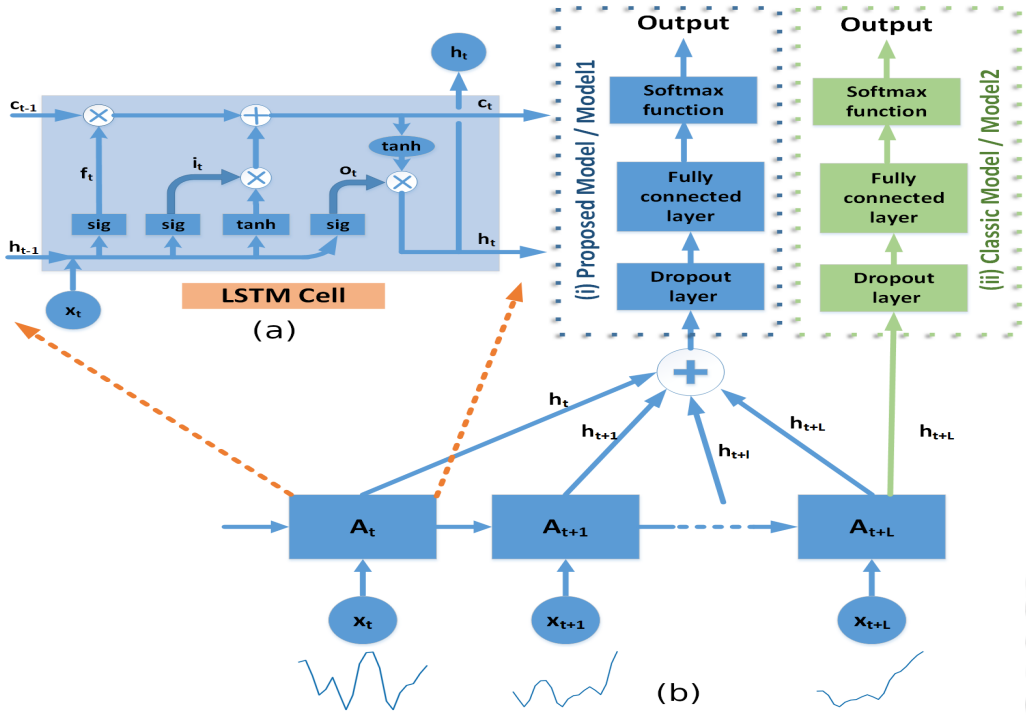

Figure 2.4: (a) The architecture of an LSTM cell. (b) Architecture of the LSTM model in the proposed framework. x_t is the input vector to an LSTM cell at time stamp t. x_{t+L} is the final input to an LSTM cell where L represents the number of segments in an ECG sequence.

different portions of an ECG signal. For this, a time duration is maintained between the last training sequence and first testing sequence. This ensures proper test protocol for the model.

Table 2.1: Number segments in one sequence for different combinations of segment length and segment shift.

Segment length (in sec)	Shift (in fraction)		
	0.25	0.5	0.75
0.1	77	39	26
0.2	37	19	13
0.4	17	9	6
0.8	7	4	3

2.1.3 LSTM Based Biometric Model

We have designed a neural network model using LSTM cells to exploit the temporal variation present in the ECG signal. LSTM is one of the many variants of the recurrent neural network (RNN) proposed in the literature. An RNN is the recurrent connection of a neuron cell, which takes past information as well as current information as its input. Thus the RNN cell utilises both the past information and the present data to make a decision. A recurrent connection of these cells utilise the temporal information

present in a time-varying signal [149]. However, the basic RNN cells suffer from exploding gradient and vanishing gradient problem. This does not allow the model to learn long-term dependencies. This motivated researchers to develop different variants of RNN cells that can account for the long term dependencies, e.g., Skip-RNN [150], LSTM [149], GRU [149], Stochastic RNN [151], Mogrifier LSTM [152], etc. Out of these, LSTM is widely used to learn representation because of its better performance [149]. The fundamental idea behind the LSTM cell is the multiple nonlinear gating units, which controls the flow of the information into as well as out of the LSTM cells, and a memory unit that retains learned representation over a period of time.

The architecture of an LSTM cell is shown in Figure 2.4 (a). The input to the LSTM cell at a time stamp t are the data (x_t), cell state (c_{t-1}), and hidden state (h_{t-1}). c_{t-1} and h_{t-1} are the output of the LSTM cell in the previous time stamp. There are three gating units in an LSTM cell, i.e., forget gate (f_t), input gate (i_t), and output gate (o_t) to control the flow of information. All the gates take x_t and h_{t-1} as input and use sigmoid function as an activation function. The weights (W) and biases (b) used in different gates are the learnable parameters of an LSTM cell. The forget gate erases the information present in c_{t-1}, which mayn't be useful further. The forget gate (f_t) is expressed as;

$$f_t = sigm(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \tag{2.1}$$

The representation learned from the current time stamp is given by $\tilde{c_t}$, which is computed using eq (2.2). Then, a point wise multiplication is done only to retain the necessary information. The input gate (i_t) and the cell state (c_t) are computed by using eq (2.3) and eq (2.4), respectively.

$$\tilde{c_t} = tanh(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \tag{2.2}$$

$$i_t = sigm(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{2.3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c_t} \tag{2.4}$$

The output of the LSTM cell at the current time stamp is the hidden state (h_t), which is used in decision making. The hidden state is obtained by removing certain information that may not contribute to the decision at time stamp t. This is done by using the output gate (o_t). The mathematical

expression to obtain o_t and h_t are shown in eq (2.5) and eq (2.6).

$$o_t = sigm(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{2.5}$$

$$h_t = o_t \odot c_t \tag{2.6}$$

In general, the output of LSTM at the final time-stamp is used for recognition purposes, which is called vanilla LSTM. This is shown in Fig 2.4 (b) as model2. However, recently it was observed that the output of the final time-stamp may not represent all the past information of the data [153]. Also, during back-propagation, the error due to first time-stamp may become negligible. So in this work, instead of taking the hidden variable only at the output, we have considered the output at each time-stamp. The proposed architecture of the LSTM model is shown in Figure2.4 (b) as model1. Here, x_t is the input to the LSTM cell, which is one ECG segment in our case. L signifies the number of ECG segments in one ECG sequence and in turn the number of LSTM cells. The output of the LSTM cell at each time-stamp is summed and passed through a dropout layer. Then the output of the dropout layer is passed through a fully connected layer. Finally, a softmax function is used to find out the probability of an ECG sequence belonging to a person. The person for which this probability is maximum is selected as the candidate person.

2.2 Experimental Results and Discussion

We have used four publicly available databases for person identification; (i) ECG-ID database, (ii) PTB database, (iii) MIT-BIH arrhythmia database, (iv) CYBHi database. ECG-ID database contains healthy ECG recordings that are recorded in an on-the-person recording setup [154]. It contains at least two recordings taken from 90 subjects. PTB and MIT-BIH arrhythmia database contains diseased data [155, 156]. MIT-BIH arrhythmia database contains 2 channels recordings taken from 47 subjects in ambulatory condition. In our study, we have used ECG data of channel 1. PTB database is a large dataset that contains recordings from 290 subjects. It contains 12 lead ECG data recorded from both healthy and diseased subjects. ECG signal of lead II is used in this study. CYBHi database was developed for ECG based biometric applications [10]. It was recorded in an off-the-person recording setup. It has two types of recordings taken from different groups of people in different recording setups; (i) long-term recordings (ii) short-term recordings. In this study, the long-term recording is used, which contains ECG acquisitions of 63 persons.

Training of the model is done using 22 ECG sequences extracted from each person. For testing of the model, 6 ECG sequences are taken from each subject. One ECG recording is used for the extraction of training and testing sequences of a person, except the ECG-ID database. In the case of the ECG-ID database, two recordings are used for ECG sequence extraction. From each recording, half of the training and testing sequences are extracted. We have also evaluated our model for inter-session data of the ECG-ID database. For this, we have extracted all the enrollment data from one session and test data from another session. In the testing mode, the result is obtained for each testing sequence. The accuracy of the model is obtained by calculating the percentage of ECG sequences classified correctly.

We have used single layer LSTM model for this application. The dimension of the hidden vector is set at 200. The input dimension varies along with the length of ECG segments. The probability of dropout of the dropout layer is set at 0.55. As the output of the fully connected layer is in probability, we have used cross-entropy loss to calculate the loss of the model. A stochastic gradient descent based network is used for optimisation of the weights of the network. The initial learning rate of the model is set at 0.01, and it is varied up to 0.001 using a cosine annealing learning rate scheduler. All the implementations of this work are done in a python environment. The neural network based model is implemented using the PyTorch environment [157]. Training and testing of the model are done using the Tesla V100 DGX GPU server of NVIDIA.

2.2.1 Results and Discussion

The proposed method for person identification is evaluated using four databases. We also experimented by varying the window size T_w and fraction of window shift to analyze the importance of the length of the ECG segment on the model's ability to capture the temporal variation. The results obtained for different datasets are tabulated in Table 2.2. For PTB dataset highest accuracy obtained is 97.3%. Similarly, for the ECG-ID database and MIT-BIH arrhythmia database, the highest accuracy obtained is 93.11% and 96.81%, respectively. Maximum accuracy of 86.24% is achieved for inter-session data of the ECG-ID database. In the case of the CYBHi dataset, an accuracy of 79.37% is obtained. From Table 2.2, it is observed that maximum accuracy for all the databases is obtained for 0.1 second segment length and 0.25 fraction of segment shift. Results obtained by varying the segment length show that higher accuracy is achieved by using smaller ECG segments. This shows that the model can capture the temporal variations better for smaller ECG segments leading to better

biometric performance. In the case of ECG signal, the temporal variation is primarily of two scales, i.e., intra-beat variation and inter-beat variation. The intra-beat variation can be captured by learning the morphological shapes of ECG complexes, i.e., P-wave, QRS complex, and T-wave. The average duration of the morphological complexes in an ECG signal is around 0.1 second (Section 1.1). This is probably the cause for LSTM based model being able to capture the intra-beat variation better for 0.1 second segment length. From Table 2.2, it is observed that, if the segment size increases then the person identification accuracy decreases. Similarly, the accuracy is observed to decrease for the increase in the fraction of the segment shift.

Table 2.2: Person identification accuracy for different combinations of segment length and segment shift. Accuracy for both the proposed LSTM model (Model1) and the vanilla LSTM model (Model2) has been tabulated for performance comparison.

Dataset	Model type	Fraction of Segment shift	Accuracy for different segment length (in percentage)			
			0.1 sec	0.2 sec	0.4 sec	0.8 sec
PTB	Model1	0.25	**97.3**	95.75	91.21	87.4
		0.5	94.77	88.56	85.46	78.22
		0.75	93.68	90.75	87.19	70.0
	Model2	0.25	**87.7**	81.84	76.95	67.59
		0.5	79.71	73.56	69.6	62.93
		0.75	78.56	78.51	68.91	52.36
MIT-BIH	Model1	0.25	**96.81**	92.55	87.23	84.40
		0.5	91.84	86.17	83.33	73.05
		0.75	92.91	88.65	76.95	58.51
	Model2	0.25	**89.36**	75.88	69.15	62.77
		0.5	75.18	60.28	61.5	51.42
		0.75	74.11	71.63	64.54	43.62
ECG-ID	Model1	0.25	**93.11**	87.52	70.02	55.87
		0.5	83.61	63.87	61.45	43.95
		0.75	79.89	70.20	48.42	35.0
	Model2	0.25	**68.9**	53.45	49.53	43.39
		0.5	48.79	38.73	39.66	32.96
		0.75	48.23	47.86	32.77	23.46
CYBHi	Model1	0.25	**79.37**	74.87	63.76	48.41
		0.5	68.52	52.66	56.88	40.47
		0.75	69.05	58.20	45.77	32.01
	Model2	0.25	**54.5**	48.18	38.09	34.39
		0.5	48.68	33.86	35.98	28.57
		0.75	34.39	42.59	36.77	24.6

In this study, we have utilized ECG sequences of 2s durations for training the biometric models. An

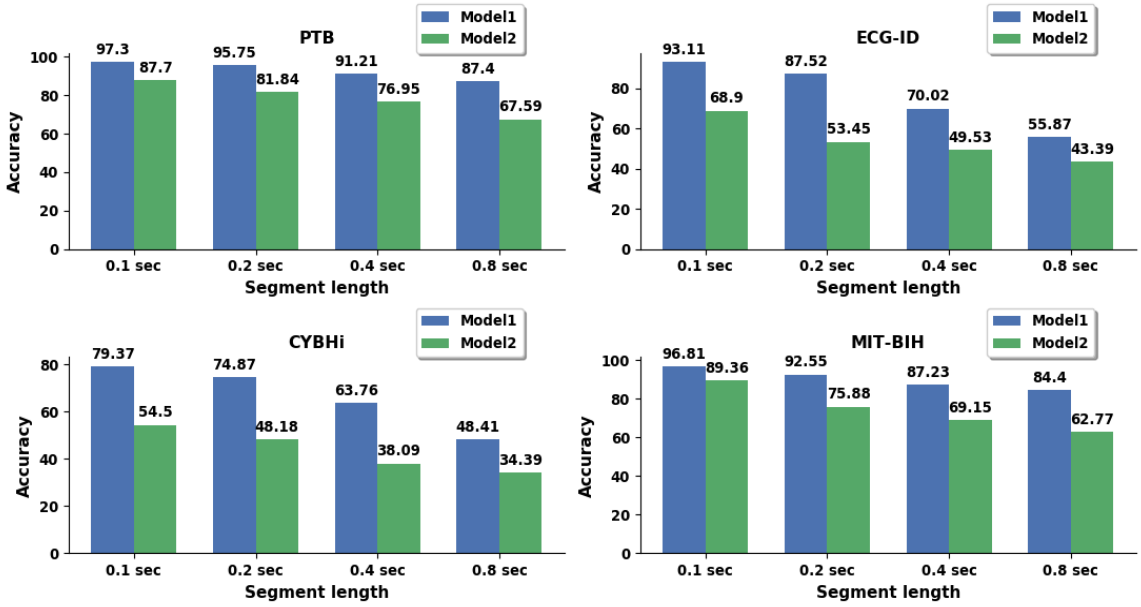

Figure 2.5: Comparison of identification accuracy between the proposed LSTM model (Model1) and vanilla LSTM model (Model2)

average heart-beat has a length of around 0.8 sec (72 beats per minute). Thus one ECG sequence contains more than one ECG beat. This approach enables the LSTM model to capture the inter-beat variation present in the ECG signal. We explored two LSTM based architectures to understand the model's capability in learning the long term temporal variation, i.e., inter-beat variation. The comparison results are tabulated in Table 2.2. Figure 2.5 shows the comparison of results for a segment shift of 0.25 fraction of segment length. From Table 2.2 and Figure 2.5, it is observed that the accuracy increases significantly by considering the output of the LSTM cell at each time-stamp (i.e. Model1). The results show that the long term temporal variation can be modeled better by the proposed Model1 compared to the vanilla LSTM model.

Our approach involved proposing a new LSTM based framework for person identification using ECG signal. Specifically, we introduced ECG segments extracted via a rectangular window as inputs to the LSTM network, enabling effective learning of the ECG signal's temporal dynamics, encompassing both intra-beat and inter-beat variations. Experimental results show that the LSTM based models can effectively learn the intra-beat variation for smaller ECG segments, i.e., 0.1s, while the inter-beat variation can be learned better by taking the output of the LSTM cell at each time-stamp. However, the proposed LSTM based model lack in effectively exploiting the multi-scale temporal representation, i.e., intra-beat variation and inter-beat variation, as evidenced by the results obtained for

the CYBHi dataset. The LSTM based model (Model1) also doesn't take into consideration that the different ECG waveform may contain different amount of biometric information corresponding to a subject.

Drawing from these concepts, we developed a novel attention based hierarchical LSTM (HLSTM) model tailored to capture the ECG signal's temporal variations in different levels of abstraction. The HLSTM is structured with two LSTM layers employing different update intervals, addressing the challenge of modeling long term temporal dependency. The incorporated attention mechanism within the HLSTM focuses on ECG complexes that carry richer biometric information unique to each individual. These ECG complexes are given more weight to learn a better biometric representation.

2.3 Hierarchical LSTM (HLSTM) Model for Person Identification and Verification

One of the major advantages of the deep neural network is its ability to learn the representation at different levels of abstraction [158]. The deep learning architectures like CNN learn the hierarchical representation in the spatial domain. Unlike CNN, the vanilla LSTM doesn't learn the hierarchical representation in the temporal dimension. To address this issue, we have designed a HLSTM model by stacking layers of LSTM. The proposed hierarchical model is designed to learn the hierarchical temporal representation, which in turn can address the long term dependency issue of the vanilla LSTM model. Figure 2.6(a) shows the proposed HLSTM model.

The layer1 of the architecture, which is represented as L_s in Figure 2.6, models the segment level representations of the ECG signal. The input to L_s are the ECG segments of an ECG sequence, which are represented by $\{x_1^s, x_2^s, .., x_t^s, ..., x_{T^s}^s\}$. Here, $x_t^s \in \mathbb{R}^{d_i}$ is the input to L_s at time stamp t with dimension d_i. The length of the input ECG sequence is T^s, which in turn is the length of time steps in L_s. $O_{t^s}^s \in \mathbb{R}^{d_s}$ is the output of L_s at time stamp t.

$$O_{t^s}^s = L_s(h_{t^s-1}^s, x_{t^s}^s) \tag{2.7}$$

The layer2 of the architecture is represented as L_B. This layer is introduced to learn the relative variation between different ECG waveforms and segments. The input to L_B is the output of L_s taken in an interval of l. The range of time steps t^B of the layer L_B depends on the interval l. The time

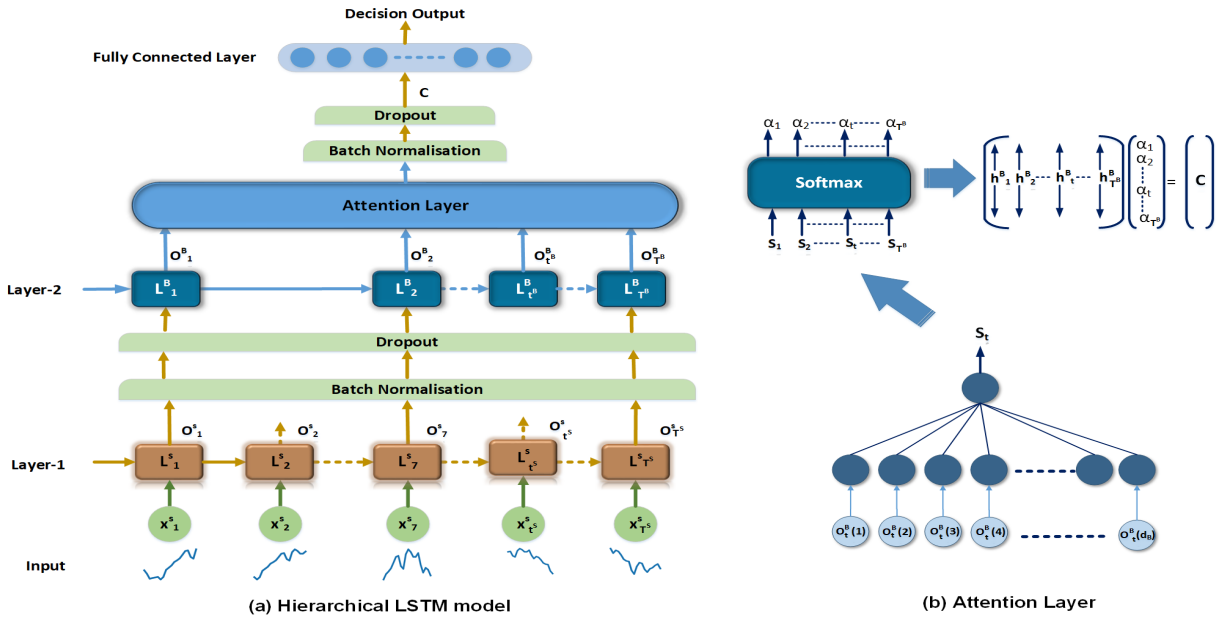

Figure 2.6: (a) Architecture of the proposed attention based hierarchical LSTM model. (b) Attention module used in this framework

steps t^B, and input $x^B_{t^B} \in \mathbb{R}^{d_s}$ of the layer L_B are expressed in eq (2.8) and eq (2.9), respectively.

$$t^B \in \{0, 1, ... \left[\frac{T^s}{l}\right] + 1\} \tag{2.8}$$

$$x^B_{t^B} \in \{O^{L_s}_1, ..., O^{L_s}_{t^B \times l}, ..., O^{L_s}_{T^{(s)}}\} \tag{2.9}$$

The output $O^B_{t^B} \in \mathbb{R}^{d_B}$ of the layer L_B at each time step t is given by;

$$O^{L_B}_{t^B} = L_B(h^{L_B}_{t^B-1}, x^B_{t^B}) \tag{2.10}$$

The output of L_B layer is given to the attention block. The attention block decides the weight that is to be given to the output vector of L_B at each time step.

2.3.1 Attention Module

The contribution of each ECG waveform towards constructing a robust biometric representation is not equal [16]. The contribution made by an ECG waveform may also vary across different subjects. So, we have introduced an attention layer to aggregate the learned representation of the HLSTM model discriminatively. The attention module can also adaptively aggregate the representations corresponding to each subject.

Various attention mechanisms have been proposed in the literature addressing the location of the

attention and weights given to the area of attention [106, 153, 159, 160]. In this work, we have used a self-attention based mechanism. Figure 2.6(b) shows the structure of the attention layer. The inputs to the attention module are the outputs of L_B layer, i.e., $\{O_0^{L_B}, O_1^{L_B}, ..., O_{T^B}^{L_B}\}$. Each output of L_B layer is passed through a fully connected layer to obtain a score value as in eq (2.11).

$$\text{score}(O_t^{L_B}) = W_A O_t^{L_B} + b_A \tag{2.11}$$

where $W_A \in \mathbb{R}^{d_B \times 1}$ and $b_A \in \mathbb{R}^1$ are the learnable weights and bias parameters of the attention layer. All the score values are fed to the softmax function to obtain the attention weights for each output vector $O_t^{L_B}$. Finally, the context vector $C \in \mathbb{R}^{d_B}$ is obtained by a weighted summation of the outputs ($O_t^{L_B}$) of L_B layer. This is expressed in eq (2.12) and eq (2.13), respectively.

$$\alpha_t = \text{softmax}(\text{score}(O_t^{L_B})) = \frac{\exp(\text{score}(O_t^{L_B}))}{\sum_{t=1}^{T^{(B)}} \exp(\text{score}(O_t^{L_B}))} \tag{2.12}$$

$$C = \sum_{t=1}^{t=T^{(B)}} c_t = \sum_{t=1}^{t=T^{(B)}} \alpha_t O_t^{L_B} \tag{2.13}$$

Subsequently, the context vector C is passed through a batch normalization layer and a dropout layer to obtain a biometric representative vector. This vector is further used either for person identification or verification, depending on the mode of operation.

2.3.2 Identification Mode

In the identification mode, the context vector is used to identify the individual from a pool of registered individuals. For the identification, the context vector is fed to a fully connected layer with a softmax activation function and an output dimension equal to the number of registered individuals. The fully connected layer gives the posterior probability output ($\hat{P}(S_k/C)$) of the context vector (C) belonging to a subject S_k.

$$\hat{P}(S_k/C) = softmax(WC + b) \tag{2.14}$$

The subject corresponding to the maximum posterior probability is identified as the candidate subject. The parameters of the model are trained by using the multiclass cross entropy loss L. The cross entropy loss L is computed over all the training samples as;

$$L = -\sum_{n=1}^{N}\sum_{k=1}^{K} P(S_k/C)log(\hat{P}(S_k/C)) \tag{2.15}$$

where N is the number of samples, and K is the number of subjects enrolled. $P(S_k/C)$ stands for the true posterior probability of the training sample belonging to subject S_k.

2.3.3 Verification Mode

In the verification mode, the context vector is used to verify the authenticity of the claimed identity. In this case, the model is first trained in the identification mode with the ECG data taken from a pool of subjects. Then the trained model is used to enrol a different set of subjects. The enrollment is done by finding out the ECG identity vector ($iECG$) for each subject in the enrolment set. The identity vector is obtained by taking the average of the normalised context vectors, C_{n_e}, for all the templates corresponding to a subject. The expression to compute $iECG$ is given in eq (2.16).

$$iECG = -\frac{1}{N_e}\sum_{n_e=1}^{N_e}\frac{C_{n_e}}{\|C_{n_e}\|} \tag{2.16}$$

During verification, the cosine similarity value between the claimant $iECG$ and the claimed $iECG$ is computed. Finally, the authenticity is verified by comparing it with a pre-set threshold value.

2.4 Experimental Results and Discussion

The proposed model is evaluated both for person identification and verification purpose. For the evaluation, we have used five publicly available databases. Out of the five databases, three are recorded in the on-the-person setting, i.e., PTB database [156], ECG-ID database [154], and two are recorded in the off-the-person setting, i.e., CYBHi database [10], UofTDB database [29]. We have evaluated our method for two scenarios; intra-session and inter-session. In the case of person identification in an intra-session scenario, both the training and the testing samples are collected from the same recording. However, unlike the works proposed in the literature, we have extracted the training and testing samples from different locations of the recording. This is done by extracting the training samples first, and then after a time duration, the testing samples are collected. This process minimises the biasing in the testing data, which otherwise would have resulted in false improvement

in the performance. For verification in the intra-session scenario, the recordings are first randomly divided into two sets without any overlap of subjects; training set and enrolment set. The training set recordings are used to train the model. The data from the enrolment set are used to obtain the enrolment data and the testing data for each subject. This process of extracting the enrolment and the testing data is similar to the training and the testing data extraction process in the identification mode.

In inter-session scenario, the training and the testing samples are collected from recordings of different sessions for identification. In case of verification in an inter-session scenario, both the recordings of a subject belonging to the training set are used for training the model. The enrolment data and the testing data are extracted from the recordings of different sessions. We have briefly described the datasets used in this work below.

2.4.1 PTB database

The PTB ECG database is a publicly available database recorded using both the conventional 12-lead ECG recording system and the frank lead recording system. The database contains records from 290 subjects. In this work, we have used the ECG signal of limb lead II, which is generally in the direction of the heart's electrical position.

2.4.2 ECG-ID database

The ECG-ID database contains data recorded from 90 subjects. Each subject in the database, except subject number 74, has at least two recordings taken in multiple different sessions. The recordings contain limb lead I ECG signals recorded for a duration of 20 seconds. In the case of the intra-session scenario, we have extracted half of the total training and testing data from one recording. While, in the case of the inter-session scenario, all the training data are extracted from one recording and the testing data from the other recording.

2.4.3 CYBHi database

The CYBHi database is a publicly available ECG database recorded in the off-the-person recording setup. The database contains ECG recordings taken from a subject in two different sessions. Depending on the duration between the two different sessions, the data have been divided into two categories; short-term and long-term. The short-term recordings are taken from an individual within

two days, and the long-term recordings are taken from an individual with an interval of three months. The long-term data have been collected from 63 subjects, and the short-term data have recordings from 65 subjects. In this work, we have used long-term data. While extracting the ECG sequences, some ECG sequences that are completely distorted have been replaced by better ECG sequences.

2.4.4 UofTDB database

The UofTDB database is the largest publicly available ECG database for the biometric application. The database contains ECG recordings of 1019 subjects recorded in the off-the-person recording setup. The data were collected for six different sessions in five different body postures. In the intra-session analysis, we have used the session-I data of 1019 subjects. For inter-session analysis, recordings of 82 subjects are used who have more than one recordings. The session-I data constitute one set of data, while another set of data consists of one recording per subject selected from the rest of the sessions. We have replaced some of the ECG sequences that are completely distorted during recording with better ECG sequences from that recording.

2.4.5 CPSC database

The CPSC ECG database is a publicly available 12-lead ECG database containing 6877 records. It has recordings whose duration varies from 6 to 60 seconds. In this work, we have considered 772 ECG records with a minimum recording length of 25 seconds. We have used the ECG signal of limb lead II for the biometric application.

2.4.6 Performance Measure

The performance of the algorithm is evaluated using five parameters, i.e., equal error rate (EER), accuracy (Acc), F1 score, Kappa score, and area under the ROC curve (AUC). EER is the trade-off point between false acceptance rate (FAR) and false rejection rate (FRR), where both FAR and FRR becomes equal. FAR stands for the rate of false acceptance of an imposter claim, and FRR stands for the rate of false rejection of a legitimate claim. In this case, all the enrolled subjects, except the legitimate ones, have been used as an imposter. The accuracy, F1 score, and Kappa score are evaluated in the case of person identification. These parameters are obtained by evaluating the number of testing templates correctly identified or falsely rejected against a total number of testing

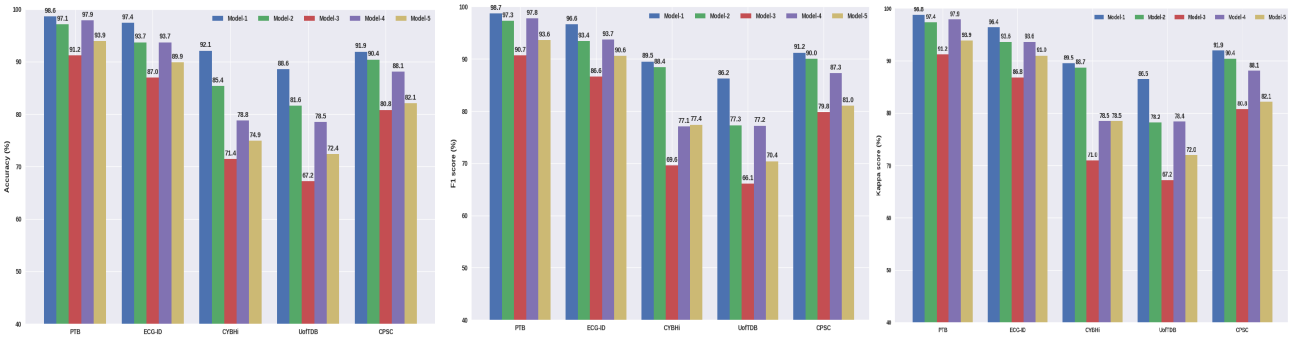

Figure 2.7: Comparison of the identification performance using five LSTM based architectures. Model-1: Attention based HLSTM model, Model-2: HLSTM model, Model-3: bidirectional LSTM model, Model-4: double layer vanilla LSTM model, Model-5: single layer vanilla LSTM model.

templates. We have also found out the threshold free AUC, which measures the model's ability to classify. The AUC is obtained for both identification as well as verification mode of operation.

2.4.7 Network Architecture

An ECG sequence containing 77 ($T_s = 77$) segments is given as input to the HLSTM model. The segment level representation vectors of dimension 150 ($d_s = 150$) are learned by layer L_s from the ECG segments of dimension 20 ($d_i = 20$). These vectors are passed through a batch normalisation (BN) layerd and a dropout layer (dropout ratio = 0.4) subsequently. Then, these vectors are fed to the L_B layer to learn a higher level representation of dimension 100 ($d_B = 100$). Finally, the context vector (C) of dimension 100 is obtained by taking an attentive weighted summation. This context vector is again passed through a BN layer and a drpout layer (dropout ratio = 0.4) to obtain a biometric representative vector corresponding to the given ECG sequence. This representative vector is further processed according to the mode of operation.

The parameters of the proposed attention based HLSTM model were trained in an end-to-end manner using a stochastic gradient descent (SGD) based optimisation algorithm. The initial learning rate of the model is set at 0.01. The learning rate is varied up to 0.001 using a cosine learning rate scheduler. The model was trained for 250 epochs with a batch size of 100. The deep learning model was implemented using the PyTorch programming library. The rest of the implementations have been done using the python programming language. All the experiments are conducted in a Titan P-100 GPU server facility.

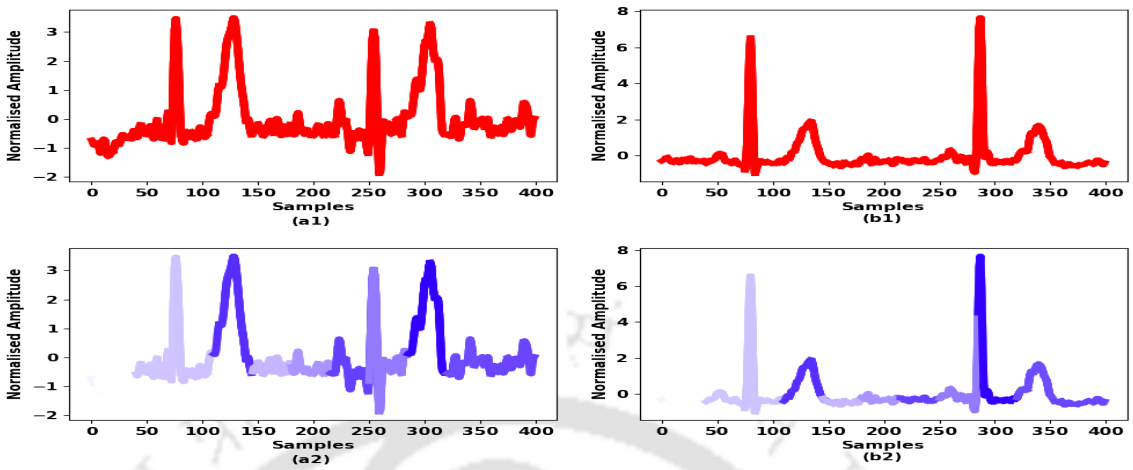

Figure 2.8: (a1) and (b1) shows the ECG sequence of two different persons. (a2) and (b2) shows the attention weight given by the model to different portions of the ECG sequence in (a1) and (b1), respectively. The portions with more colour saturation correspond to more attention weights.

Table 2.3: Identification Performance of the Proposed Model (Intra-Session Scenario)

	PTB	ECG-ID	CYBHi	UofTDB	CPSC
ACC	98.6	97.4	92.1	88.6	91.9
F1	98.7	96.6	89.5	86.2	91.2
Kappa	98.8	96.4	89.5	86.5	91.9

2.4.8 Results and Discussion

The identification performance of the proposed HSLTM model for all the five datasets are tabulated in Table 2.3. From Table 2.3, it can be observed that the HLSTM model performs well for both *off-the-person* and *on-the-person* ECG data. We have compared the performance of the proposed HLSTM model (Model-1) with the HLSTM model without attention (Model-2), bidirectional LSTM (Model-3), vanilla LSTM with two layers (Model-4) [149] and the single layer vanilla LSTM model (Model-5) [149]. The performance comparison is shown in Figure 2.7. It can be observed that the one layer vanilla LSTM model and the bidirectional LSTM model performs very poorly for the person identification application. From Figure 2.7, it can be observed that the HLSTM model (Model-2) performs significantly better than the vanilla LSTM model. The performance of the HLSTM model is also superior to the double layer vanilla LSTM model in the case of *off-the-person* ECG data. This shows that the HLSTM model can capture the temporal variation of the ECG signal in different abstractions by exploiting the multiscale temporal information of the ECG signal. Among the five models, the proposed attention

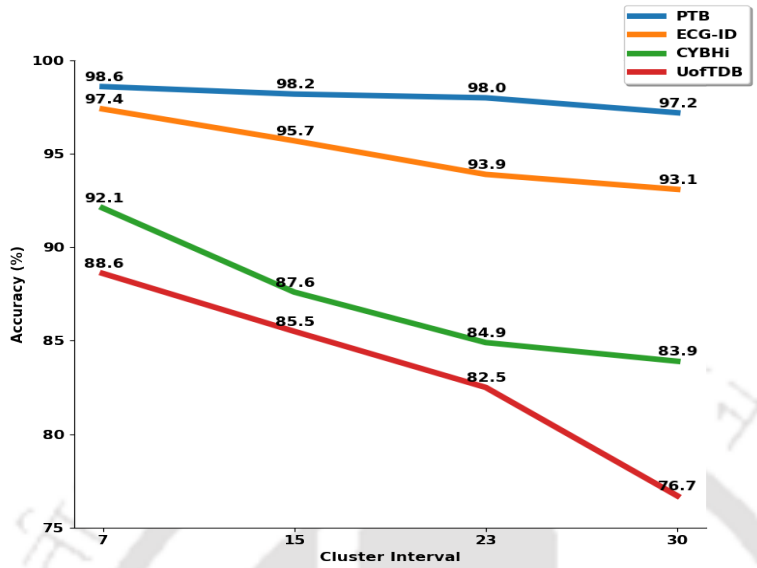

Figure 2.9: Variation of person identification accuracy of the model for different interval lengths.

based HLSTM model performs the best. The proposed model's attention mechanism helps in giving more weight to the segments, which are important for the identification problem. Figure 2.8 shows the attention weight given to ECG sequences taken from two different persons. Figure 2.8(a1) and Figure 2.8(b1) are the ECG sequences of two different persons and their respective figures with normalised attention weights are given in Figure 2.8(a2) and Fig 2.8(b2). The portions with more colour saturation have been given more attention weight compared to the portions with less colour saturation. From Figure 2.8(a2), it can be observed that more weight has been given to the T wave. In the case of ECG sequence in Figure 2.8(b2), more weight has been given to both the QRS complex and T wave. Thus, the attention mechanism learns to give more weight to the portions of the ECG signal that have more biometric information corresponding to each subject.

The identification accuracy of the proposed model for different interval lengths (l) has been calculated, and the results are shown in Figure 2.9. From Figure 2.9, it can be observed that the best performance is obtained for an interval of 7 and the identification accuracy decreases as we increase the interval length. This may be because smaller interval lengths give better temporal resolution, and thus, the model can learn the multiscale temoral information better for smaller interval length.

The EER and AUC values of the proposed model operated in the identification mode are obtained for different databases. The results obtained are tabulated in Table 2.4. The AUC obtained in all the cases is close to 1. This shows that the trained model is good at the identification task. We have

Table 2.4: EER and AUC value obtained by the model in the Identification mode

Database		PTB	ECG-ID	CYBHi	UofTDB	CPSC
Intra session	EER (in %)	0.89	1.70	3.98	2.19	1.96
	AUC	0.9996	0.9995	0.9979	0.9983	0.9993
Inter session	EER (in %)	-	3.16	13.43	11.36	-
	AUC	-	0.9905	0.9391	0.9423	-

obtained a significantly low EER of 0.89%, 1.7%, and 1.96% for the intra-session analysis of PTB, ECG-ID, and CPSC ECG databases, respectively. Similar performance has also been observed in the case of the CYBHi and UofTDB databases with a minimum EER of 3.98% and 2.19%, respectively. The results of the inter-session analysis have also been tabulated in Table 2.4. A low EER of 3.16% has been obtained for the ECG-ID database, which is of the on-the-person recording category. The inter-session analysis of the CYBHi dataset has been done in two ways. Once the model is trained using (S1) data and then evaluated using (S2) data, which is represented by (S1-S2). The next time the model is trained using (S2) data and evaluated using (S1) data, which is represented by (S2-S1). We have obtained an EER of 13.43% and 12.05% for the S1-S2 session and S2-S1 session analysis, respectively. The respective AUC score obtained for the S1-S2 session and S2-S1 session are 0.939 and 0.951. In the case inter-session analysis using the UofTDB database, we have obtained an EER of 11.36%.

Table 2.5: EER and AUC value obtained by the model in Verification mode

Database	Percentage of data used for training					
	40%		60%		80%	
	EER (in %)	AUC	EER (in %)	AUC	EER (in %)	AUC
PTB	1.89	0.9974	1.39	0.9991	1.33	0.9987
ECG-ID	4.81	0.991	4.2	0.9921	3.13	0.9946
UofTDB	1.94	0.9973	2.10	0.9971	2.06	0.9968
CPSC	3.45	0.9916	2.81	0.9945	2.80	0.997

The performance of the model in the verification mode has been shown in Table 2.5. The evaluation is done by using different percentages of data as training data and the rest as enrolment data. The lowest EER is obtained when more data are used for training. However, it can be observed

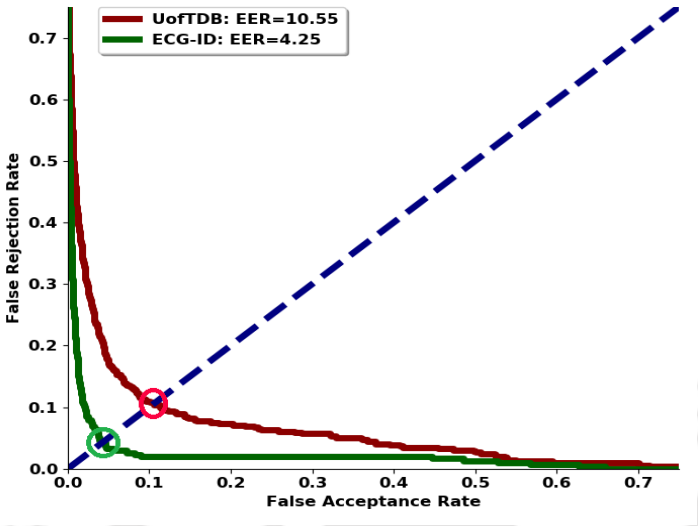

Figure 2.10: DET curve obtained from the inter-session analysis of ECG-ID and UofTDB database in the verification mode.

that the model gives comparable results with less training data also. This shows that the performance of the model is robust. The model has also been evaluated in the verification mode using the inter-session data. Figure 2.10 shows the graph between FRR and FAR of the results obtained for the inter-session analysis using UofTDB and ECG-ID database. We have obtained an AUC of 0.95 and 0.983 for the UofTDB and ECG-ID database, respectively. An EER of 4.25% and 10.55% are obtained for inter-session analysis of the ECG-ID database and UofTDB databases, respectively.

2.4.9 Comparison

A comparison of the proposed model with the state-of-the-art works has been made in this section. Various aspects of the biometric system have been taken into account for comparison. We have considered recent deep learning based methods, i.e., [26, 28, 31, 40, 69, 70], for comparison in both on-the-person category and off-the-person category. A comparison of the literature works on the on-the-person ECG database has been made in Table 2.6.

From Table 2.6, it can be observed that the proposed method performs best in the on-the-person category. Unlike most of the methods in the literature, we have evaluated our model for all the subjects present in the database. For better comparison with the models with less enrolment in the case of the PTB database, we have evaluated our model for 52 healthy subjects and 100 subjects taken randomly. Our model achieves 100% accuracy for both the cases outperforming the existing

Table 2.6: Comparison with the existing works using on-the-person ECG data. FPD stands for fiducial point detection.

Database	Works	Session type	No. of Subjects	FPD	Amount of training data	Acc.
PTB	[26] (2019)	intra	52	Yes	40 s	100%
	[27] (2013)	intra	100	Yes	100 templates of 2-4 s	99.48%
	[28] (2020)	intra	290	No	96 s on average	94.9%
	[42] (2019)	intra	285	Yes	15 ECG beats	97.19%
	Proposed	intra	290	No	12.5s	98.6%
	Proposed	intra	100	No	12.5s	100%
	Proposed	intra	52	No	12.5s	100%
ECG-ID	[69] (2018)	intra	50	Yes	36 s on average	96.63%
	[161] (2015)	intra	90	Yes	-	83.88%
	Proposed	intra	90	No	12.5s	97.4%
	Proposed	inter	90	No	12.5s	92.32%

models. From the table, it can be observed that the proposed model requires the least amount of data for training. The non-fiducial point based CNN architecture proposed in [28] requires an average of 96 seconds of training data per subject, which is significantly more than the 12.5 seconds of training data used in this work. The LSTM based method proposed in [40] is implemented in our experimental setup, and the results are compared in Table 2.6.

Table 2.7 represents the comparison of the proposed model with the works in literature for the off-the-person ECG database. From Table 2.7, it can be observed that the proposed model performs better than the methods proposed in the literature except in two cases, i.e., inter-session (S1-S2) analysis of [31] and inter-session analysis of UofTDB [30]. However, the proposed model performs better than [31] in the (S2-S1) session analysis. Similarly, our method performs significantly better than the work in [30] for the intra-session analysis of UofTDB data. The literature methods benefit from the cosine similarity based outlier removal method employed in the preprocessing stage. Thus, the works in literature have used similar templates of a subject for the training and testing.

From the comparative analysis, it can be understood that the proposed biometric system has several important advantages over the existing methods. The attention based hierarchical LSTM model requires a very low duration of enrolment data, i.e., 12.5 seconds of ECG recording. Similarly,

Table 2.7: Comparison with the existing works using off-the-person ECG data. FPD stands for fiducial point detection.

Database	Works	Session type	No. of Subjects	FPD	Amount of training data	EER
CYBHi	[31] (2018)	intra	63	Yes	50% of data	1.33%
	[31] (2018)	inter (S1-S2)	63	Yes	50% of data	12.78%
	[31] (2018)	inter (S2-S1)	63	Yes	50% of data	13.93%
	[70] (2019)	intra	63	Yes	90% of data	4.47%
	Proposed	intra	63	No	12.5s	3.98%
	Proposed	inter (S1-S2)	61	No	12.5s	13.43%
	Proposed	inter (S2-S1)	61	No	12.5s	12.05%
UofTDB	[31] (2019)	inter	82	Yes	50% of data	14.27%
	[30] (2016)	intra	1012	Yes	80% of data	7.89%
	[30] (2016)	inter	82	Yes	80% of data	10.10%
	Proposed	intra	1019	No	12.5s	2.19%
	Proposed	inter	82	No	12.5s	11.36%

the model requires only 2 seconds of data during verification. The proposed framework follows a non-fiducial point based approach, which makes it more suitable for application in a practical scenario. The LSTM based framework learns the representation of the temporal variation of the ECG signal in two different levels of abstraction. This might allow the model to learn the differences between different cardiac signals, which arise due to the physiological differences of cardiovascular systems [3]. The physiological differences between different cardiovascular systems are reflected in the systolic and diastolic phase of the ECG signal. The proposed model is designed to exploit these differences optimally with the help of the attention mechanism. The attention mechanism could give more weight to portions of the ECG signal important for identification purpose, which leads to better performance of the model.

2.5 Summary

This chapter introduces a novel LSTM based framework for ECG based person identification and verification. First, we proposed to use the ECG segments extracted using rectangular window as inputs to the LSTM based model, which enables the biometric model in effectively learning the temporal variation of the ECG signal. We investigated the effect of different segment lengths on model's capability to learn the temporal representation. Experimental results showed that the LSTM based models can effectively learn the intra-beat variation for smaller ECG segments.

Further, we designed an attention-based HLSTM model aimed at capturing multi-scale temporal representations for biometric applications. The proposed HLSTM model is composed of two layers of stacked LSTM network with different update intervals. this hierarchical architecture efficiently learned the ECG signal's temporal representation in different abstraction, which also addresses the long-term dependency problem. It was observed that a smaller interval length within the hierarchical structure yielded better results in learning the hierarchical representation of the ECG signal. The attention mechanism embedded within the model could identify crucial ECG complexes specific to each subject, assigning them higher weightage and thereby enhancing the model's performance

The major advantage of the proposed method is that it doesn't require the detection of any fiducial points. The proposed framework exhibited superior performance, utilizing minimal enrollment data—merely 12.5 seconds for each subject and a testing data duration of 2 seconds. Compared to the state-of-the-art model, our proposed framework performed notably better for both on-the-person and off-the-person ECG data. Overall, this model showcases considerable potential for practical application of ECG-based biometric system.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

3

Enhanced Morphological and Multiscale Temporal Learning for Biometric Representation

Contents

The electrocardiogram signal presents distinctive morphological traits unique to each individual, making it a suitable candidate for biometric applications. Given that the ECG signal captures the individual's unique electro-physiological processes over time, extracting its temporal representation becomes vital for an effective biometric system. In the previous chapter, we introduced an ECG-based biometric system, focused on learning the temporal representation. An essential aspect of the proposed framework is its independence from the detection of fiducial points, ensuring the system's consistent performance in real-world applications. Another aspect of a robust biometric system is its permanence, which signifies the ability of the extracted biometric features to remain sufficiently invariant over a given period.

The differences in the morphological characteristics in the ECG signal of each subject manifest primarily in the shape and the temporal dynamics of the four basic ECG waveforms, i.e. P wave, QRS complex, ST segment and T wave (Figure 3.1) [16, 23]. The variation in the morphological shape of the ECG waveform across different subjects is a major biometric cue. The shape and duration of a waveform also vary for different ECG waveforms. From Figure 3.1, it can be observed that the shape and duration of the P wave, QRS complex, and T wave are different. Thus, the biometric information can be better modeled by learning multi-scale morphological representation. Apart from the morphology of the ECG waveforms, the temporal variation in the ECG presents crucial biometric information. The temporal variation in ECG signal is broadly of two scales, i.e. intra-beat variation and inter-beat variation [107]. The intra-beat variation signifies the relative variations of the ECG waveforms within an ECG beat, while the inter-beat variation signifies the beat-to-beat variation. The multi-scale local morphological shape of the ECG waveform and multi-scale temporal variation encompass the person specific information.

As discussed in the Section 1.3, the existing works lack in exploiting the temporal variation of the ECG signal explicitly for biometric application. To address this, we presented the HLSTM model in Chapter 2, which learns the temporal variation of the ECG signal in different abstractions. The HLSTM model based framework employs a blind windowing method for extracting ECG sequences. However, the LSTM cells of HLSTM model are tasked with learning both the local morphological representations and the signal's temporal dynamics, potentially limiting the model's representational capacity. Additionally, the blind windowing approach introduces a challenge by including the baseline ECG signals that contains minimal biometric information. Although, an attention layer is introduced

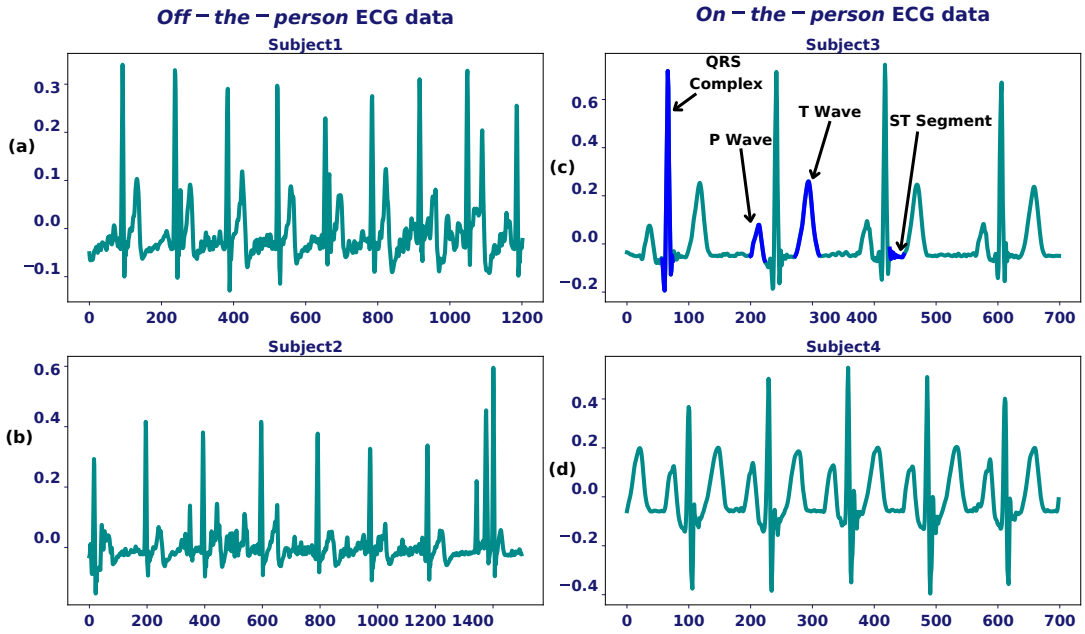

Figure 3.1: Filtered ECG Signals from Four Subjects: (a) and (b) display ECG signals acquired using the *off-the-person* setup, while (c) and (d) show *on-the-person* ECG recordings. Notably, variations in signal shape, duration, and temporal dynamics are evident among the subjects. Furthermore, it is noteworthy that *off-the-person* ECG records exhibit pronounced high-frequency noise artifacts post noise removal.

to emphasize the significant biometric waveforms, its effectiveness might be compromised due to the presence of noise and multiple ECG beats. Thus, the existing biometric methods lack in the following aspects: 1) Effectively capturing the multi-scale morphological information and temporal dynamics of ECG signals. 2) Enhancing biometric representation by selectively leveraging informative ECG morphology. 3) Extensive multi-session analysis of non-fiducial point based approach for *off-the-person* ECG record.

This chapter introduces the multi-scale temporal dynamics learning network (MSTDLNet), a novel approach for capturing both the local morphological representation and multi-scale temporal dynamics of ECG signals for biometric applications. Unlike the existing works that learn the multi-scale temporal representation [162–166], we leverage the fine-to-coarse flow of information within a stacked convolutional network to learn the multi-scale temporal representation. Specifically, we have designed a convolutional kernel based multi-scale enhanced morphological representation learning (MSE-MRL) module to learn the ECG waveform's morphological representation better. Further, the multi-scale temporal dynamics of the ECG signal are learned by innovatively integrating two layers of LSTM networks at different hierarchical levels. The architecture of the proposed MSTDLNet is

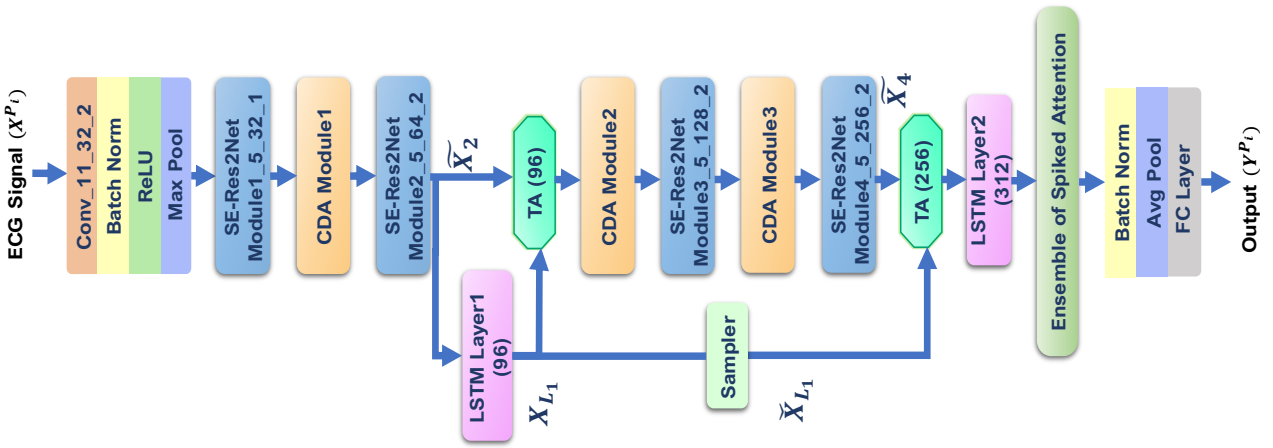

Figure 3.2: Architecture of the MSTDLNet. The convolutional parameters are denoted as Conv_(kernel size)_(number of filters)_(stride). The SE-Res2Net module's parameters are denoted as SE-Res2Net Module(layer number)_(kernel size)_(number of filters)_(stride). Padding for convolution operation is int(kernel_size/2). The LSTM network's output dimension is indicated in bracket. Maxpooling specifications are: window size = 3, stride = 1, and padding = 1.

depicted in Figure 3.2. Our proposed framework offers a comprehensive approach for learning both local morphological representation and temporal dynamics of ECG signals at multiple scales.

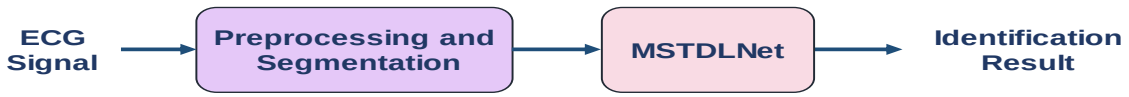

Figure 3.3: Block diagram of the proposed ECG based biometric system

3.1 MSTDLNet Based Biometric System

The ECG based biometric identification process is illustrated in Figure 3.3. It involves ECG acquisition, noise removal and segmentation, representation learning, and identification. We begin by formally defining the identification problem and introducing basic notations. The significance of the composing modules of MSTDLNet is described subsequently.

3.1.1 Problem Formulation

In this work, we pose the biometric identification problem as a time-series classification task. The biometric dataset is denoted as $D = \{(X^{P_i}, Y^{P_i})\}$, where $X^{P_i} =< x(1), x(2), .., x(t), .., x(T) >$ is an ECG biometric template, consisting of T sample points, extracted from subject P_i. The scalar $x(t)$ signifies a single sample value within the ECG template vector X^{P_i}. The label $Y^{P_i} \in \mathbb{R}^N$ is a one-hot vector

assigned to person P_i, where $P_i \in \{1, 2, ..N\}$ and N stands for the number of subjects in the dataset D.

3.1.2 Preprocessing and Segmentation

The ECG signal gets corrupted with various high-frequency and low-frequency noises during acquisition. Therefore, we first filter the ECG signal using a band-pass butter-worth filter with a higher cut-off frequency of 40 Hz and a lower cut-off frequency of 0.5 Hz. Following that, we employed the wavelet based filtering process and morphological filters, akin to the methodology outlined in Chapter 2 of our work, to remove noise within the pass band. Then, the ECG signal is resampled to $200Hz$. We have followed a reference-free blind segmentation process for extracting the training and testing biometric templates. Specifically, we employ a rectangular window with a duration of 2 seconds and 25% shift to extract the training templates. The testing templates are obtained using the same rectangular window with 100% shift. In this work, we have extracted 22 training templates amounting to 12.5 s of data and 6 testing templates amounting to 12 s of data from each subject P_i.

3.1.3 Proposed MSTDLNet

The proposed MSTDLNet is composed of the CNN based MSE-MRL module and two layers of LSTM network. The architecture of the MSTDLNet model is depicted in Figure 3.2. The MSE-MRL module is proposed to effectively learn the enhanced multi-scale morphological representation of the ECG signal. It consists of four layers of interleaved scale enhanced Res2Net (SE-Res2Net) module and the committee of dual attention (CDA) module. The SE-Res2Net module is designed to learn the multi-scale morphological representation $\tilde{X}_l$ while enhancing representation from biometric-rich scale. Subsequently, the learned representation is passed through the CDA module to enhance the person specific information encoded in $\tilde{X}_l$. This is achieved by giving more attention to specific ECG waveforms through a novel ensemble of spiked attention (ESA) module and selectively emphasizing features carrying substantial biometric information, as well. These processes are expressed using eq. 3.1 and eq. 3.2

$$\tilde{X}_l = \textit{SE-Res2Net}(X_l) \tag{3.1}$$

$$X_{l+1} = CDA(\tilde{X}_l) \tag{3.2}$$

The lower layers of the CNN architecture (i.e. MSE-MRL module) learn the low-level information with high temporal resolution, while the upper layers capture the contextual information with lower temporal resolution. In the proposed framework, we have attempted to exploit this hierarchical representation to learn the multi-scale temporal representation. This is achieved by feeding the features obtained from different layers of MSE-MRL module to the LSTM networks. The output feature map, $\tilde{X}_2 =< \tilde{X}_2(1), .., \tilde{X}_2(t_2), .., \tilde{X}_2(T_2) >$, from SE-Res2Net Module2 is fed as input to the 1^{st} layer of LSTM network (L_1). Here, $\tilde{X}(t_2) \in \mathbb{R}^{d_2}$ is a representation vector of $\tilde{X}_2$ at instance t_2. T_2 stands for total number of time instances in $\tilde{X}_2$. This is expressed in Eq. 3.3.

$$X_{L_1}(t_2) = L_1(h_{L_1}(t_2 - 1), \tilde{X}_2(t_2)) \tag{3.3}$$

Here, $X_{L_1}(t_2) \in \mathbb{R}^{d_{L_1}}$ is the output of L_1, at instance t_2. $h_{L_1}(t_2 - 1) \in \mathbb{R}^{d_{L_1}}$ is the hidden vector of L_1 at instance $t_2 - 1$ and d_{L_1} is the output dimension of L_1. Considering the sampling frequency of input ECG template (X^{P_i}) at 200 Hz, the representation vectors fed to L_1 has a frequency of 50 Hz. Therefore, it can be hypothesized that the LSTM network, L_1, models the temporal variation within an ECG waveform. Subsequently, the temporal variation information is innovatively infused to the MSE-MRL module by aggregating X_{L_1} and $\tilde{X}_2$ using a novel temporal aggregation (TA) module. Finally, the multi-scale temporal representation is learned by feeding the aggregated representation of $X_{Agg4} \in \mathbb{R}^{d_{Agg4}}$ as input to 2^{nd} layer of LSTM network (L_2). X_{Agg4} is obtained by aggregating $\tilde{X}_4$ and sampled output of X_{L_1}. The vectors in X_{L_1} are sampled to match total time instances in $\tilde{X}_4$, i.e., T_4 (Eq. 3.4). Eq. 3.5 shows the relation between T_2 and T_4. The vectors in X_{Agg4} has a frequency of 12.5 Hz. Thus, L_2 models the inter-waveform temporal variation of the ECG signal, which is in a different scale than L_1. The multi-scale temporal representation learning process is expressed in eq 3.6 and eq 3.7.

$$\breve{X}_{L_1} = Sampler(X_{L_1}) \tag{3.4}$$

$$=< X_{L_1}(1), X_{L_1}(4), ., X_{L_1}(8), ., X_{L_1}(4(T_4 - 1)) >$$

$$T_4 = [\frac{T_2}{4}] \tag{3.5}$$

$$X_{Agg4} = TA(\breve{X}_{L_1}, \tilde{X}_4) \tag{3.6}$$

$$X_{L_2}(t_4) = L_2(h_4(t_4 - 1), X_{Agg4}(t_4)) \tag{3.7}$$

The composing modules of MSTDLNet are described in subsequent subsections.

3.1.3.1 Scale Enhanced Res2Net (SE-Res2Net) Module

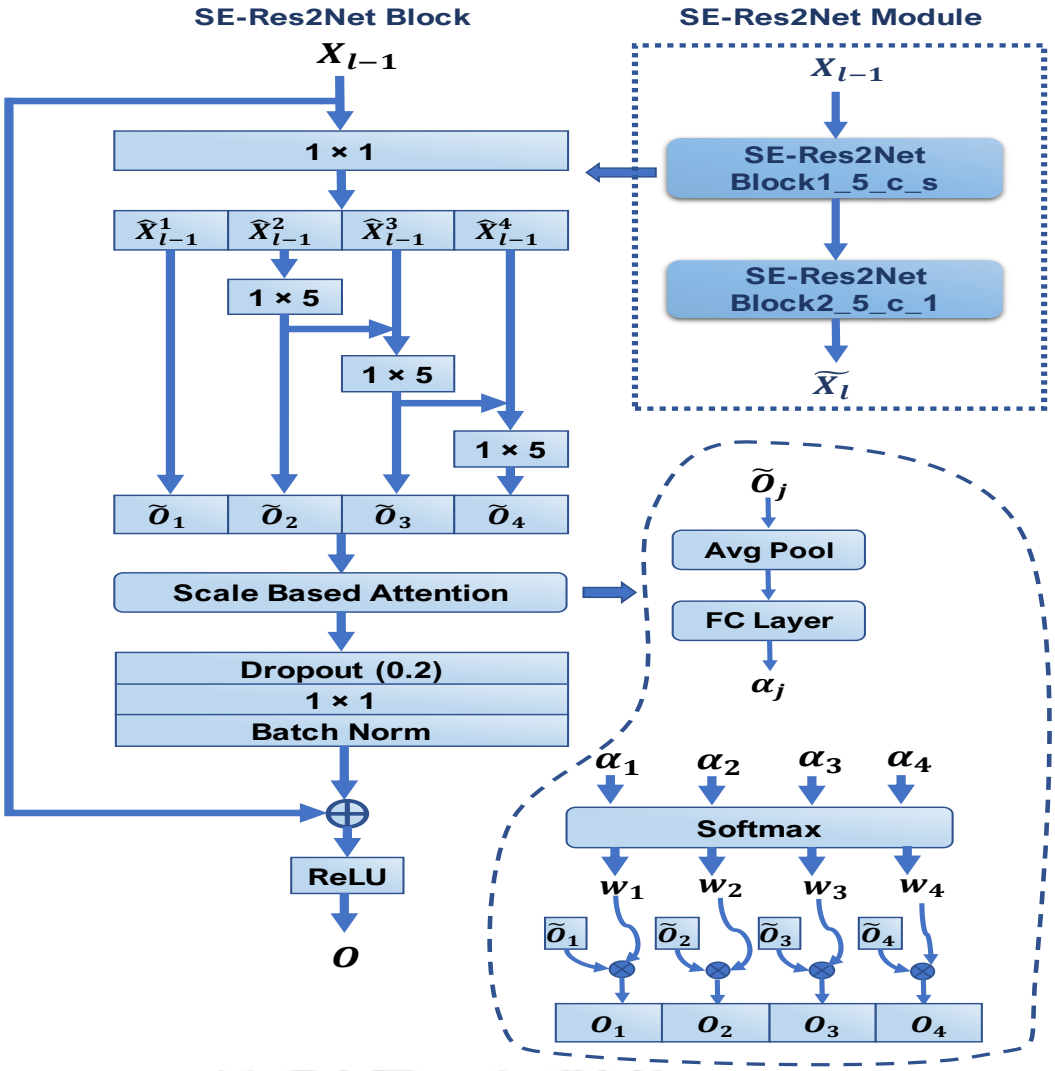

Figure 3.4: Architecture of the SE-Res2Net module.

The SE-Res2Net module is one of the basic building block of the proposed MSTDLNet. Figure 3.4 shows the architecture of the SE-Res2Net module that consists of two layers of SE-Res2Net blocks. The SE-Res2Net block is designed to enhance representation of specific scales that contain crucial biometric information. This is accomplished by incorporating a scale-based attention block at the output of the Res2Net block. The Res2Net block learns the multi-scale representation by utilizing multiple receptive fields within a residual block [167]. It involves connecting a group of filters in a hierarchical residual like structure (Figure 3.4). The incoming feature map $X_{l-1} \in R^{d_{l-1}}$ is first

convolved using a 1×1 kernel followed by splitting it into S subset feature maps, denoted as $\hat{X}_{l-1}^{j}$. Here, $j \in \{1, 2.., S\}$ and $\hat{X}_{l-1}^{j} \in R^{d/S}$. Except $\hat{X}_{l-1}^{1}$, each $\hat{X}_{l-1}^{j}$ are convolved with a respective 1×5 kernel denoting K^j. The output $\tilde{O}^j$ can be expressed as in eq 3.8.

$$\tilde{O}^j = \begin{cases} \tilde{O}^j & \text{if } j = 1 \\ K^j(\hat{X}_{l-1}^{j}) & \text{if } j = 2 \\ K^j(\hat{X}_{l-1}^{j} + \tilde{O}^{j-1}) & \text{if } 2 < j \leq s \end{cases} \tag{3.8}$$

The kernel K^j receives the output of possibly all the kernel $K^n \forall n < j$ and feature subset $\hat{X}_{l-1}^{j}$ as input. Thus, the Res2Net block learns multi-scale representation using multiple receptive fields. Finally, the output subset feature maps, $\tilde{O}^j, j \in \{1, 2.., S\}$, are fed to an attention block to augment specific feature map subsets. The attention process is expressed using eq 3.9, eq 3.10, and eq 3.11.

$$\alpha_j = W_{l-1}(AvgPool(\tilde{O}^j)) + b_{l-1} \tag{3.9}$$

$$w_j = \text{softmax}(\alpha_j) = \frac{\exp(\alpha_j)}{\sum_{j=1}^{S} \exp(\alpha_j)} \tag{3.10}$$

$$O^j = w_j \times \tilde{O}^j \tag{3.11}$$

The scale based attention process involves squeezing of the subset feature maps $\tilde{O}^j$ followed by a fully connected network. Here, $W_{l-1} \in \mathbb{R}^{d/S \times 1}$ is learnable parameter and α_j is a scalar output. Subsequently, the attention weights w_j are obtained using the softmax activation function and multiplied with $\tilde{O}^j$ to obtain the scale enhanced multi-scale representation.

3.1.3.2 Committee of Dual Attention (CDA) Module

The CDA module consists of two parallel attention networks: (i) Ensemble of Spiked Attention (ESA) and (ii) Contextual Attention (CA). It is strategically crafted to augment the discriminative features within the biometric representation obtained from the SE-Res2Net module. Figure 3.5 depicts the architecture of the proposed CDA module.

$$X_l = S_A(\tilde{X}_l) + C_A(\tilde{X}_l) \tag{3.12}$$

The ESA network enhances the contribution of ECG waveforms with significant biometric infor-

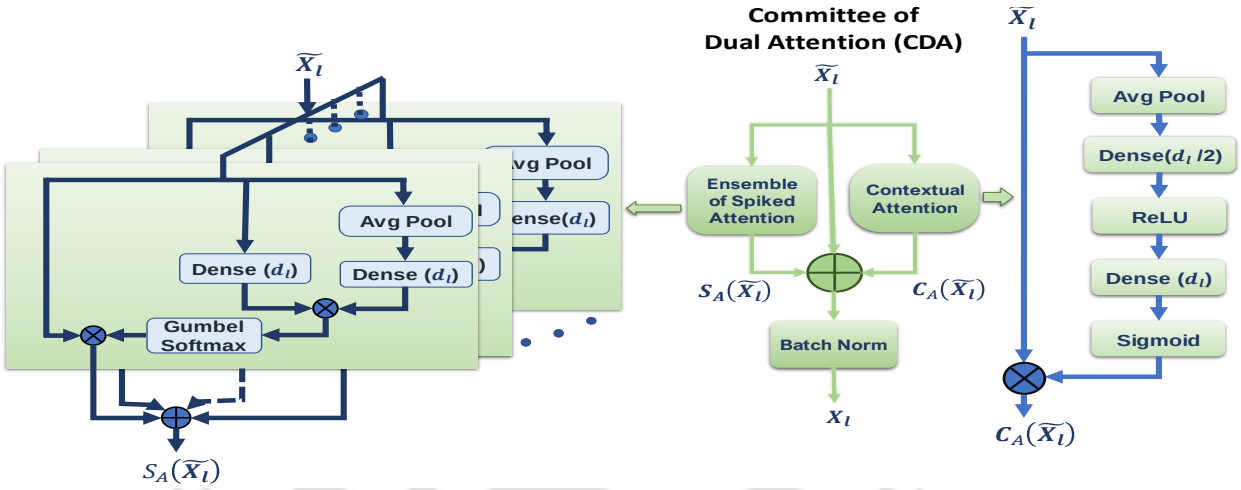

Figure 3.5: Architecture of the committee of dual attention (CDA) module. (a) Architecture of the ESA module. (b) Architecture of CA module.

mation. The biometric information contained in different ECG waveforms varies across subjects [16]. Therefore, it becomes crucial to enhance the representation of informative ECG waveforms. Unlike conventional attention mechanisms, spiked attention (SA) selects a single representative vector rich in biometric details. This allows the model to emphasize specific ECG waveforms among many in the temporal dimension. This is achieved by employing a Gumbel softmax function. To address the potential presence of multiple biometric rich ECG waveforms within a 2s ECG template, we propose an ensemble of spiked attentions connected in parallel. This setup facilitates the selection of biometrically significant vectors from multiple temporal locations. The SA mechanism involves first obtaining a query vector $H_q \in \mathbb{R}^{d_l}$ in the latent space through an average pool followed by linear transformation operation on $\tilde{X}_l \in \mathbb{R}^{d_l}$ (eq 3.13). Here, $W_q \in \mathbb{R}^{d_l \times d_l}$, and $b_q \in \mathbb{R}^{d_l}$ are learnable parameters. Then the matrix of key vectors, $H = [H_1, ..H_{t_l}.., H_{T_l}]$ is obtained using eq 3.14, where $H \in \mathbb{R}^{d_l}, W \in \mathbb{R}^{d_l \times d_l}$, and $b \in \mathbb{R}^{d_l}$. Subsequently, the correlation factor r_{t_l} and attention weight a_{t_l} corresponding to each key vector are obtained using eq 3.15 and eq 3.16, respectively. Here, $r_{t_l} \in \mathbb{R}$ and $a_{t_l} \in \{0, 1\}$. The Gumbel-Softmax sampling is expressed in eq 3.16. Here, g_0, g_1 are the i.i.d. samples drawn from Gumbel(0,1) and π_0, π_1 are the Bernoulli distribution of attention mask $A = [a_1, .., a_{t_l}, ..a_{T_l}]$. The softmax function is employed along with Gumbel-Max trick to produce a continuous differentiable approximation of the attention mask [168, 169]. When the temperature $\tau \to 0$, the Gumbel-Softmax distribution becomes identical to the Bernoulli distribution. Finally, the attention output is obtained using eq 3.17. Here, N is the number of SA blocks employed in the ESA module and $a^n_{t_l}$ is the attention

value of n^{th} SA block.

$$H_q = W_q(AvgPool(\tilde{X}_l)) + b_q \tag{3.13}$$

$$H_t = W(\tilde{X}_l(t_l)) + b \tag{3.14}$$

$$[r_1, .., r_{t_l}, ..r_{T_l}] = H_q^T H \tag{3.15}$$

$$[a_1, .., a_{t_l}, ..a_{T_l}] = \frac{exp((log(\pi_1) + g_1)/\tau)}{\sum_{j\in 0,1} exp((log(\pi_j) + g_j)/\tau)} \tag{3.16}$$

$$S_A(\tilde{X}_l(t_l)) = \sum_{n=1}^{N} a_{t_l}^n \tilde{X}_l(t_l) \quad \text{for } t_l \in \{1, .., T_l\} \tag{3.17}$$

$$U_l = \frac{1}{T_l} \sum_{t_l=1}^{T_l} \tilde{X}_l(t_l) \tag{3.18}$$

$$C_A(\tilde{X}_l) = \sigma(W_2 \times (ReLU(W_1 \times (U_l))) \times \tilde{X}_l \tag{3.19}$$

The features learned across the temporal dimension may encode various attributes, i.e., shape, duration, elevation angle etc., related to different ECG waveform. The CA mechanism is employed to enhance the discriminating biometric features learned by the network [**?**, 170]. The CA mechanism involves squeezing operation which generates a global representative vector followed by an excitation network that recalibrates the input feature maps using a set of attention weights. Figure 3.5 (b) depicts the architecture of the CA module. First, the contextual biometric information is embedded in a representative vector by global average pooling of the local descriptors as expressed in eq 3.18. The vector $U_l \in \mathbb{R}^{d_l}$ represents the channel wise statistics of the local biometric descriptors. Subsequently, the complex channel-wise inter-relationship is encoded through a non-linear transformation using a bottleneck layer. The bottleneck layer consists of two fully-connected layers around a ReLu non-linearity. $W_1 \in \mathbb{R}^{d_l \times d/2}$ and $W_2 \in \mathbb{R}^{d_l/2 \times d}$ are the parameters of the fully connected layer. Finally, the relevant biometric features are highlighted by multiplying the input feature maps with the attention weights. Eq 3.18 and eq 3.19 describes the operation of the CA module.

3.1.3.3 Temporal Aggregation (TA) Module

The TA module is introduced to effectively fuse the morphological representation from the MSE-MRL module and the temporal variation from the LSTM network. Figure 3.6 shows the architecture of the TA module. The aggregation process involves concatenation of inputs from both modules followed by

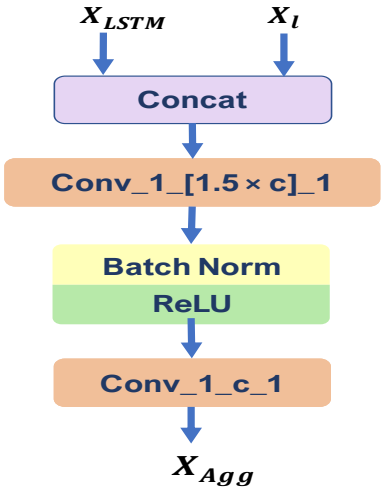

Figure 3.6: Architecture of the TA module.

a depth convolution using a 1×1 kernel. The output undergoes normalization and a ReLU non-linear activation, resulting in reduced dimension representation in a latent space. Finally, the aggregated output is obtained by passing it through another layer of depth convolution. The TA module generates a reduced dimension representation by encoding the complex relation between the morphological and temporal features via nonlinear transformation, potentially eliminating redundant information in both representations.

3.2 Experiments

3.2.1 Datasets

The proposed MSTDLNet model is assessed using three publicly available ECG biometric databases, i.e., ECG-ID database [154], CYBHi database [10], and UofTDB database [29]. The performance is evaluated for two scenarios, i.e., intra-session analysis and inter-session analysis. In the intra-session analysis, both training and testing data are extracted from the same session recording. Unlike prior works [23], our approach draws training and testing data from different segments of the recording to minimize bias on model's performance. This is achieved by initially extracting training data and subsequently collecting testing data with a time delay. In the inter-session analysis, the training and testing data are collected from different recordings taken across multiple sessions of a subject. We have also conducted experiments on ECG recordings of UofTDB dataset acquired in different physical postures. A succinct database overview follows.

3.2.1.1 ECG-ID Database

The ECG-ID database contains 310 lead I ECG recordings taken from 90 subjects. The database contains multiple ECG recordings sourced from each of the 89 subjects (except subject number 74) in multiple different sessions. Each recording contains 20 s ECG data sampled at 500 Hz. In this study, we have selected two recordings from each subject for analysis.

3.2.1.2 CYBHi Database

The CYBHi database contains ECG recrodings that are collected in an *off-the-person* recording setup. The database contains two ECG recordings per person collected in two different sessions. The ECG data of the CYBHi dataset are of two broad categories, (i) Short-term data, (ii) Long-term data. The Short-term data contain two recordings that are collected within a span of two days from an individual, while long-term recordings span a minimum of three months. In this study we have used the long-term recordings available for 63 subjects.

3.2.1.3 UofTDB Database

The UofTDB database is the largest publicly available ECG biometric dataset. It contains recordings acquired from 1019 subjects in an *off-the-person* recording setup. The recordings were collected over six sessions in five different body-postures. This dataset is developed to study the impact of different body postures and time-gap over biometric information. Table 3.1 shows the statics of ECG recordings across different sessions and body-postures.

Table 3.1: Statistical distribution of number of subjects in different sessions

Session	Sit	Stand	Exercise	Supine	Tripod
S1	1012	0	0	0	0
S2	72	72	0	0	0
S3	76	5	71	0	0
S4	63	0	0	0	0
S5	0	0	0	63	63
S6	65	65	0	0	0

3.2.2 Evaluation Method

In this study, we have used accuracy (Acc), F1-score (F1), precision (Pre), Kappa score (Kappa), and expected error rate (EER) as the performance metric to assess the model's effectiveness in biometric identification. The EER score is derived by considering all templates, except that of the legitimate subject, as impostor. The performance metrics are calculated by evaluating the number of correctly identified test templates or falsely rejected test templates against the total number of test templates.

3.2.3 Implementation Details

The ECG templates of duration $2s$ (i.e. 400 sample points) obtained through blind segmentation process is given as input to the proposed MSTDLNet. The dimension of different blocks of the MSTDLNet network is mentioned in Figure 3.2. The L_1 LSTM's input(d_2) and hidden dimensions (d_{L_1}) are set at 64 and 96, respectively. Similarly, the L_2 LSTM operates with input (d_{Agg4}) and hidden dimensions (d_{L_2}) of 256 and 312, respectively. The ESA block incorporates 5 SA architectures. The SE-Res2Net block is configured with a scale (S) of 4 and the dropout rate is set at 0.25 at all the cases.

The network parameters are optimized using an Adam optimizer. While training the biometric model, we set an initial learning rate of 0.0001 and a weight decay of 0.0001. The learning rate is varied to $(1/5)$th of initial learning rate using a cosine learning rate scheduler. The biometric model is trained with a batch size of 16. All the deep learning models are implemented using the PyTorch framework. The experiments are conducted on an NVIDIA A100 GPU facility.

3.2.4 Baseline Model

We systematically evaluate the efficacy of the proposed MSTDLNet model against a spectrum of well-established deep learning models. Our performance analysis encompasses HLSTM [171], ResNet [172], MS-ResNet [173], and ResNeXt [174], benchmarked against the MSTDLNet model. The HLSTM model captures the multi-scale temporal information, while MS-ResNet is a multi-scale CNN based architecture. Additionally, we incorporate two well-established baseline biometric models, Heart-ID [46] and DCNN [64], drawn from the literature, to provide comprehensive comparative insights.

Table 3.2: Performance Comparison on the ECG-ID Dataset

Model	Intra Session					Inter Session				
	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER
HLSTM [171]	97.39	97.45	98.07	97.36	1.33	93.82	93.33	93.94	93.75	2.19
ResNet34 [172]	97.39	97.35	98.00	97.36	3.10	91.95	91.34	92.53	91.86	1.82
MS-ResNet [173]	96.83	96.85	97.64	96.80	3.93	94.94	94.20	95.24	94.89	1.60
ResNeXt [174]	96.65	96.74	97.44	96.61	0.99	94.19	93.22	95.08	94.13	1.29
Heart-ID [46]	63.69	62.80	66.94	63.28	10.48	56.74	56.35	60.42	56.25	11.61
DCNN [64]	92.55	92.54	93.61	92.47	7.10	84.46	83.39	85.09	84.28	9.03
Proposed	**98.70**	**98.70**	**99.03**	**98.68**	**0.55**	**96.44**	**96.07**	**96.22**	**96.40**	**0.95**

3.2.5 Results on ECG-ID dataset

The biometric identification performance of the proposed MSTDLNet model and baseline models is tabulated in Table 3.2 for two scenarios, i.e., intra-session analysis, and inter-session analysis. In intra-session analysis half of the training and testing data are collected from each of the recording of a subject. Table 3.2 shows that the MSTDLNet model consistently outperforms the baseline model across all performance metrics by a significant margin. The MSTDLNet model achieves an identification accuracy of 98.70% for intra-session analysis and 96.44% for inter-session analysis. Similarly, an EER of 0.55% and 0.95% is obtained for intra-session and inter-session analysis, respectively. The promising results, specifically for the inter-session analysis, holds a great promise for a robust ECG based biometric system. Table 3.2 also demonstrates that the MS-ResNet architecture performs notably better than the ResNet architecture for inter-session analysis. The HLSTM model, designed for learning multi-scale temporal representation, performs similarly to the MS-ResNet architecture, outperforming the basic ResNet architecture. This observation underscores the significance of both multi-scale morphological and temporal representations in learning crucial biometric information that remain invariant over time. These findings are further reinforced by the robust performance of the proposed MSTDLNet model in multi-session analysis.

3.2.6 Results on CYBHi dataset

Table 3.5 presents a performance comparison of our proposed model against baseline models using the CYBHi dataset. The intra-session performance is assessed for both S1 and S2 sessions of CYBHi

dataset, while the inter-session performance is examined once with S1 and S2 sessions serving as training sets once (S1-S2 and S2-S1). The results from Table 3.5 clearly demonstrate the marked performance improvement of our MSTDLNet across all performance metrics. Notably, the MSTDLNet achieves accuracy and F1-score improvements exceeding 8% compared to the next best-performing models, HLSTM and MS-ResNet, in intra-session (S1) analysis. Similarly, the EER value has decreased by over three times compared to MS-ResNet model. In the case of inter-session analysis (S1-S2), the MSTDLNet showcases a 8.4% accuracy boost and a 7.92% F1-score enhancement over the MS-ResNet model. Correspondingly, the EER value displays a substantial 3.08% improvement compared to the MS-ResNet model. Notably, the relative performance improvement of MSTDLNet over baseline models is more pronounced in inter-session analysis. These findings underscore the significance of effectively modeling the multi-scale local morphological and temporal features of ECG signal in enhancing biometric representation. Importantly, these representations enable effective generalization to recordings from different sessions, a critical attribute for a robust biometric system.

3.2.7 Results on UofTDB dataset

Table 3.3: Performance Comparison on the UofTDB dataset for The Sitting Posture

Model	Intra Session					Inter Session (Sit-Sit)				
	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER
HLSTM [171]	88.21	87.74	89.69	88.20	1.80	52.85	47.82	51.58	52.26	**10.97**
ResNet34 [172]	86.67	86.14	87.87	86.66	16.04	47.76	43.29	46.19	47.12	19.50
MS-ResNet [173]	88.16	87.70	89.40	88.15	16.08	54.67	49.66	52.11	54.12	14.98
ResNeXt [174]	85.57	85.02	86.80	85.56	6.69	51.22	46.97	49.61	50.62	14.26
Heart-ID [46]	15.49	11.57	13.15	15.41	43.56	21.34	20.13	23.40	20.37	26.25
DCNN [64]	72.73	71.46	74.75	72.71	23.08	46.34	41.63	42.71	45.68	21.12
MSTDLNet	**88.73**	**88.31**	**90.09**	**88.72**	**1.67**	**59.35**	**54.75**	**57.36**	**58.85**	**11.06**

Table 3.4: Performance Comparison on the UofTDB dataset for Different Body Postures

Model	Inter Session (Sit-Stand)					Inter Session (Sit-Supine)					Inter Session (Sit-Tripod)				
	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER
HLSTM [171]	35.19	31.93	36.98	34.38	17.57	45.50	41.37	44.92	44.62	12.11	39.15	36.26	42.93	38.17	15.60
ResNet34 [172]	26.13	22.25	26.18	25.21	26.61	31.22	28.64	36.19	30.11	24.03	24.87	23.20	29.18	23.66	27.02
MS-ResNet [173]	34.77	30.63	35.49	33.96	20.55	46.30	42.44	46.16	45.43	14.17	37.57	33.24	35.15	36.56	20.80
ResNeXt [174]	34.77	30.86	36.25	33.96	17.45	44.18	39.72	45.76	43.28	14.86	34.66	32.16	36.17	33.60	25.21
Heart-ID [46]	15.84	13.96	15.40	14.79	26.91	18.78	16.79	18.22	17.47	26.56	12.43	12.11	16.45	11.02	30.03
DCNN [64]	30.45	27.94	32.85	29.58	25.75	36.77	33.12	35.73	35.75	22.54	26.19	22.66	25.72	25.00	25.34
Proposed	**42.59**	**37.78**	**41.34**	**41.87**	**16.92**	**51.85**	**47.49**	**52.50**	**51.08**	**13.52**	**47.35**	**43.42**	**50.10**	**46.51**	**17.21**

Table 3.5: Performance Comparison on the CYBHli Dataset

Model	Intra Session (S1)					Intra Session (S2)					Inter Session (S1-S2)					Inter Session (S2-S1)				
	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER
HLSTM [171]	87.30	86.60	88.55	87.10	5.41	86.24	85.95	87.81	86.02	5.13	59.52	56.92	59.50	58.87	12.22	60.32	56.74	58.86	59.68	11.66
ResNet34 [172]	83.86	83.16	86.24	83.60	4.07	82.54	81.91	84.63	82.26	4.49	53.97	51.17	55.89	53.23	12.72	56.61	53.11	56.15	55.91	13.60
MS-ResNet [173]	87.30	86.89	89.91	87.10	3.93	89.15	88.90	90.46	88.98	2.56	61.38	58.14	63.61	60.75	11.34	63.49	60.85	66.56	62.90	10.00
ResNeXt [174]	72.75	71.11	73.45	72.31	7.16	73.54	72.07	74.68	73.12	6.94	57.14	54.16	55.64	56.45	11.88	54.50	51.31	54.40	53.76	13.01
Heart-ID [46]	44.18	42.93	44.57	43.28	16.07	43.39	41.85	44.85	42.47	15.35	22.75	22.17	27.05	21.51	23.19	24.07	22.29	23.42	22.85	21.50
DCNN [64]	79.89	78.88	80.64	79.57	7.66	78.57	77.69	79.41	78.23	7.61	55.82	52.15	55.18	55.11	16.65	47.88	45.25	48.36	47.04	15.02
Proposed	**96.03**	**96.05**	**96.71**	**95.97**	**1.06**	**94.71**	**94.58**	**95.59**	**94.62**	**1.10**	**69.84**	**66.06**	**69.26**	**69.35**	**8.26**	**67.72**	**64.36**	**67.35**	**67.20**	**7.94**

The biometric identification performance, using ECG recordings from different sessions of the UofTDB database, is summarized in both Table 3.3 and Table 3.4. Table 3.3 presents the results for intra-session data taken from 1019 subjects and inter-session data comprising 82 subjects, all recorded in a sitting posture. To assess the impact of different body postures on biometric identification performance, the results are obtained on ECG recordings taken in different body postures and tabulated in Table 3.4. In all the cases, the ECG recordings of subjects in sitting posture (S1 session) is used as training data. From Table 3.3, it can be observed that the MSTDLNet gives superior performance across all the performance metrics. Notably, in inter-session analysis within sitting postures, MSTDLNet achieves an accuracy of 59.35% and an F1-score of 54.75%, which is significantly higher than the baseline models. Furthermore, it is observed that the CNN based architectures perform less effectively compared to LSTM based models in terms of EER. This might be due to the fact that the temporal variation presents crucial biometric information that are independent of extreme noise that may be present in an *off-the-person* ECG signal. Interestingly, MSTDLNet also exhibits an improvement in terms of the EER metric, possibly due to its robust learning of multi-scale temporal representation, complemented by strong multi-scale morphological features.

Table 3.4 presents results for records acquired in standing, supine, and tripod postures. MSTDLNet gives a performance improvement of 7.4%, 5.55% and 8.2% in terms of accuracy compared to the next best-performing model for standing, supine and tripod posture, respectively. In this analysis, it can also be observed that the LSTM-based models excel in terms of the EER score for different body postures. This emphasizes the significance of temporal representation in developing a robust ECG based biometric system for different practical scenarios. When comparing the results from Table 3.3 and Table 3.4, it becomes apparent that MSTDLNet's performance improvement is notably more pronounced in inter-session analysis compared to intra-session analysis. This observation underscores the permanence property of the learned multi-scale representation by MSTDLNet. This observation validates that the multi-scale representation learned by MSTDLNet exhibits permanence over time, resulting in enhanced performance during inter-session analysis.

3.2.8 Ablation Experiments on MSTDLNet

We conducted a comprehensive series of experiments to evaluate the impact of various MSTDLNet modules, and the resulting outcomes are presented in Table 3.6. To begin, we assessed the identification results using the base Res2Net module, which does not incorporate the CDA module, LSTM,

or TA module. As Table 3.6 illustrates, the performance of the basic Res2Net architecture demonstrates a moderate level of effectiveness in terms of identification. However, it's noteworthy that the Res2Net architecture outperforms the MS-ResNet architecture, highlighting its effectiveness in learning the multi-scale morphological representations. The major advantage of Res2Net architecture lies in its modular design and improved performance with fewer parameters (Res2Net($S = 4$) in Table 3.8).

To understand the effect of enhancing scale-specific information, we have obtained the results using the SE-Res2Net block in the same neural network architecture. From Table 3.6, it can be observed that the SE-Res2Net block improves performance across all the metrics on different datasets. This improvement can be attributed to the enhanced representation of biometric-rich scales, potentially minimizing the influence of noisy scales in the process. This is evident in the results obtained on the CYBHi and UofTDB datasets. Further, the impact of ESA module is studied by plugging only the CDA module without the CA block. The results obtained are tabulated in Table 3.6 under MARes2Net architecture. It can be observed that the MARes2Net model improves the identification performance almost across all the datasets. A marked performance improvement can be observed for the intra-session analysis of CYBHi dataset and inter-session analysis of UofTDB dataset. Then, the CA module is introduced in the MSE-MRL model. From Table 3.6, it can be observed that the model's performance improves for all the dataset except the UofTDB dataset. The performance improvement is nominal which may be attributed to the better representation enhancement by the ESA module.

Finally, the effectiveness of learning multi-scale temporal representation is assessed by evaluating results for the proposed MSTDLNet module and MSE-MRL module coupled with LSTM network. The MSE-MRL + LSTM network is implemented by feeding the output of MSE-MRL module to an LSTM layer. This learns the temporal variation of the ECG signal. From Table 3.6, it can be observed that learning the temporal representation significantly improves the identification performance across all the databases, including multi-session analysis. It can be observed that the temporal representation significantly improves the EER value and other metrics. This shows that the temporal representation play a crucial role in learning biometric representation that remains invariant across different sessions. Finally, the effectiveness of the multi-scale temporal representation learned by the MSTDLNet can be assessed by the results presented in Table 3.6. It can be observed that the MSTDLNet model significantly improves the identification performance, which are majorly reflected

in the inter-session analysis of CYBHi and UofTDB. The accuracy increases to 69.84% and 59.35% against 66.67% and 57.93% obtained by MSE-MRL + LSTM network for CYBHi and UofTDB datasets, respectively. Similarly, the EER value of UofTDB improves to 11.06% compared to 12.98% obtained by MSE-MRL + LSTM network. This shows that the multi-scale temporal representation learned by the proposed MSTDLNet model learns robust biometric representation for ECG signal recorded in *off-the-person* set-up. It is also noteworthy that the multi-scale temporal representation learned by MSTDLNet exhibit permanence which is crucial for effective biometric system.

3.2.9 Effect of Scale (S) in SE-Res2Net Block

The scale (S) parameter controls the multi-scale features learned by the Res2Net block. A larger scale allows for learning rich multi-scale information with a potentially larger receptive field. To assess the impact of the scale parameter, we conducted experiments using only the Res2Net block in our architecture with varying scale values. The results are shown in Figure 3.7. Notably, as depicted in Figure 3.7, the Res2Net model with $S = 1$, which does not learn multi-scale features, exhibits inferior performance compared to instances with $S = 2$, $S = 4$, and $S = 8$. Overall, performance tends to improve with an increase in the scale value. This can be attributed to the multi-scale morphological representation learned by the Res2Net model. The performance improvement for $S = 8$ is marginal compared to $S = 4$. Even $S = 4$ gives better performance than $S = 8$ in few instances, e.g., UofTDB dataset. Table 3.7 compares the number of parameters for MS-ResNet architecture and Res2Net model with different scale values. The number of parameters increases by more than two fold for increasing $S = 4$ to $S = 8$. Therefore, we have selected $S = 4$ for learning multi-scale representation in our work.

Table 3.6: Ablation Experiments on The MSTDLNet Model

Model	CYBHi (Intra Session: S1)					CYBHi (Inter Session: S1-S2)					ECG-ID (Inter Session)					UofTDB (Inter Session: Sit-Sit)				
	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER	Acc	F1	Pre	Kappa	EER
MS-ResNet	87.30	86.89	89.91	87.10	3.93	61.38	58.14	63.61	60.75	11.34	94.94	94.20	95.24	94.89	1.60	54.67	49.66	52.11	54.12	14.98
Res2Net	89.15	89.02	90.17	88.98	4.72	63.23	59.41	62.21	62.61	11.66	95.13	94.30	95.30	95.08	2.06	55.49	50.94	53.75	54.95	15.06
SE-Res2Net	90.74	90.28	91.12	90.59	3.36	63.49	59.63	63.09	62.90	10.80	95.51	95.13	96.38	95.45	2.06	56.71	51.71	53.36	56.17	14.28
MARes2Net	91.53	91.53	92.61	91.40	2.13	63.76	59.38	61.81	63.17	10.62	95.51	95.26	96.63	95.45	0.89	57.52	52.83	55.66	57.00	15.38
MSE-MRL	91.80	91.55	92.30	91.67	2.34	64.55	61.46	63.20	63.98	10.27	95.88	95.50	97.12	95.83	1.07	56.10	51.91	54.29	55.56	16.68
MSE-MRL + LSTM	94.71	94.60	95.45	94.62	1.08	66.67	63.03	67.42	66.13	**8.21**	96.07	95.30	95.87	96.02	**0.81**	57.93	53.64	55.85	57.41	12.98
MSTDLNet	**96.03**	**96.05**	**96.71**	**95.97**	**1.06**	**69.84**	**66.06**	**69.26**	**69.35**	8.26	**96.44**	**96.07**	**96.22**	**96.40**	0.95	**59.35**	**54.75**	**57.36**	**58.85**	**11.06**

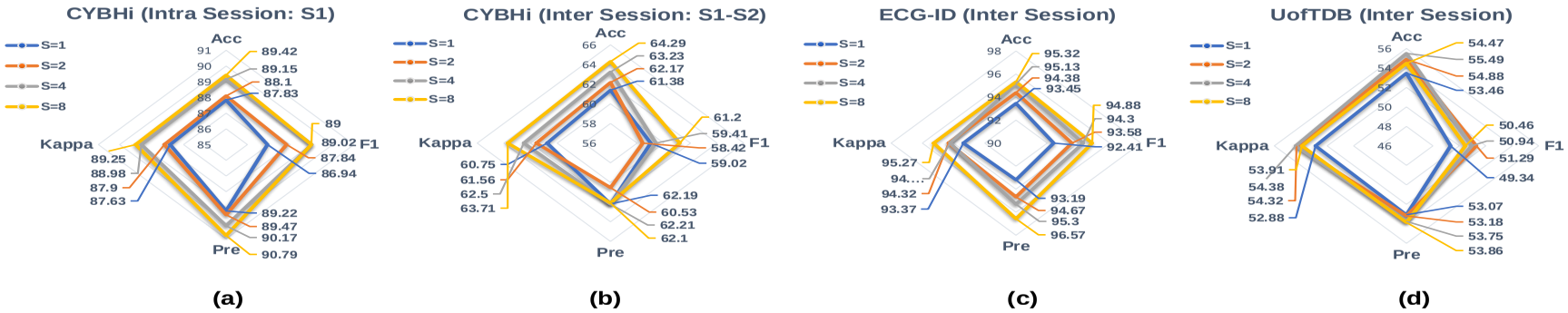

Figure 3.7: Performance Comparison of the Res2Net model for different scale values, i.e.,$S = 1$, $S = 2$, $S = 4$, and $S = 8$

Table 3.7: Comparison of Number of Model Parameters for MS-ResNet model and Res2Net model with different scale values

Models	MS-ResNet [173]	Res2Net (s = 1)	Res2Net (s = 2)	Res2Net (s = 4)	Res2Net (s = 8)
Parameters	4656754	913962	1163458	2814770	6117394

3.2.10 Effect of Ensemble of Spiked Attention (ESA) Module

The ESA module is designed to enhance representation from specific ECG waveforms. First, we obtain the results by replacing the Gumbel-Softmax function with Softmax function to assess the impact of enhancing only specific ECG waveforms. The results are presented in Figure 3.8. We can observe that the ESA module with Gumbel-Softmax perform consistently better than the Softmax based attention. This is particularly pronounced in the case of inter-session analysis of CYBHi dataset, which typically contains substantial noise. This may be attributed to the fact that softmax based attention results a distributed attention map that may not concentrate its weight on ECG waveforms that are biometrically more significant. While the SA block enhances specific ECG waveform, it also penalizes low biometric information content in the baseline. This may also help in suppressing noisy ECG waveform that might be present in an *off-the-person* ECG record. For better visualization, we have plotted the attention map of the ESA module in Figure 3.10. In Figure 3.10(a), an ECG template is presented, while Figure 3.10(b)-(e) present the corresponding attention maps. It can be observed that the ESA module attempts to give more attention to the T-wave as in Figure 3.10(d)-(f). The ESA module also appropriately attends to the QRS complex (Figure 3.10(b)) and P wave (Figure 3.10(c)).

We have obtained results for $N = 1$, $N = 3$, and $N = 5$ to assess the impact of employing multiple parallel SA blocks, as depicted in Figure 3.9. It can be observed that the performance of the MSTDLNet model improves as the number of SA blocks increase. This highlights the importance of enhancing representation from multiple ECG waveforms that are present in an ECG template of duration 2 seconds. It is also worth noting that increasing the value of N increases computation complexity. Therefore, we have set $N = 5$ in this work.

3.2.11 Model Parameters

The number of trainable parameters used in the proposed MSTDLNet model and baseline models

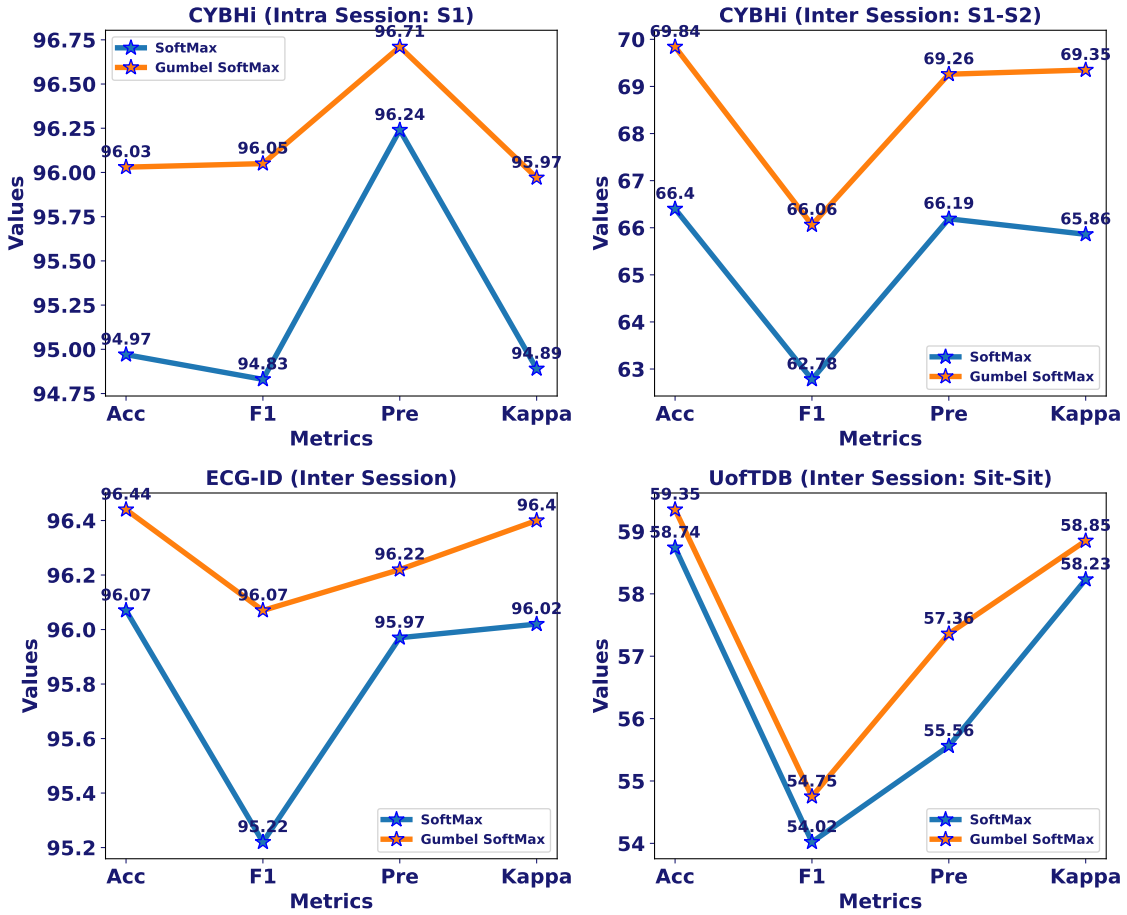

Figure 3.8: Comparison of MSTDLNet Model's performance with Gumbel-Softmax function and Softmax function

is presented in Table 3.8. The Heart-ID model has least number of model parameters, i.e., 137194 followed by the HLSTM model with 212883 parameters. However, the HLSTM model perform significantly better than the Heart-ID model. The performance of HLSTM model is comparable to ResNet34 which uses 3022578 parameters. The MSTDLNet model has 5098214 number of parameters which is comparable to MS-ResNet and DCNN model, and notably less than the ResNeXt architecture. However, the performance of the MSTDLNet is significantly more than the baseline models.

Table 3.8: Comparison of Number of Model Parameters

Models	HLSTM [175]	Resnet34 [172]	MS-ResNet [173]	ResNeXt [115]	Heart-ID [46]	DCNN [64]	MSTDLNet
Parameters	212883	3022578	4656754	22363794	137194	5152812	5098214

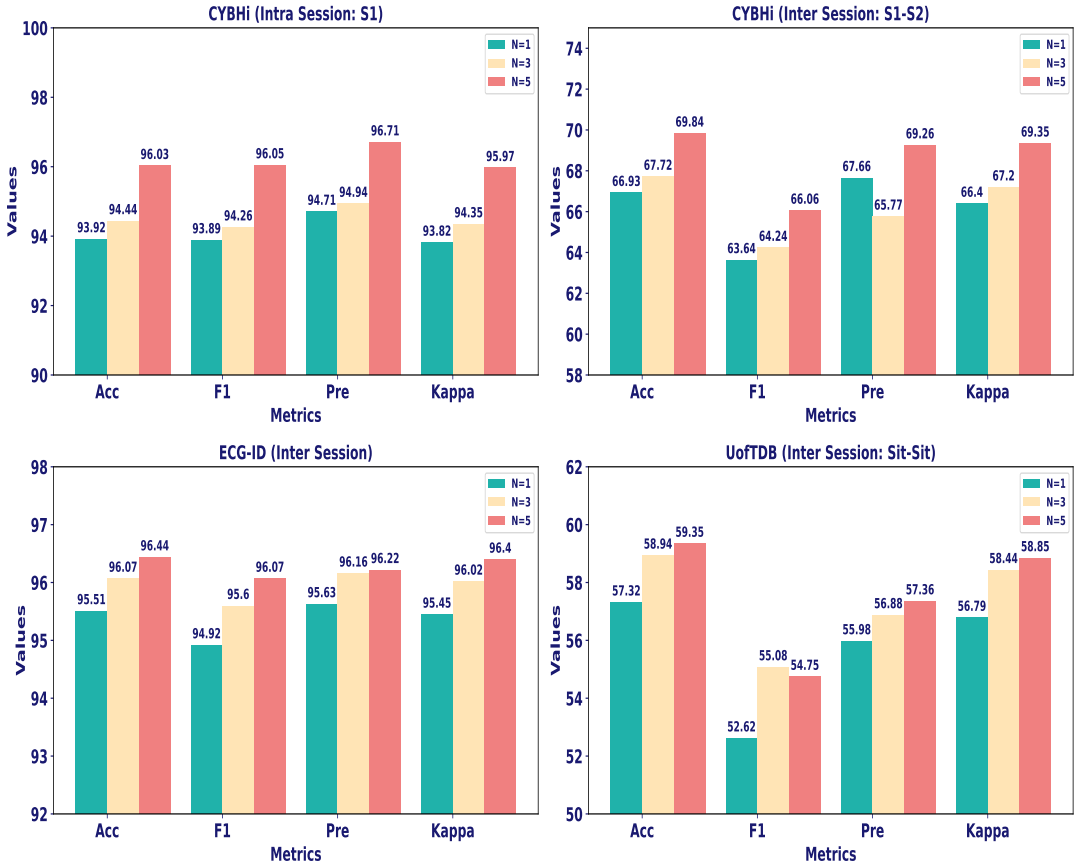

Figure 3.9: Comparison of MSTDLNet model's performance for different values of N in ESA

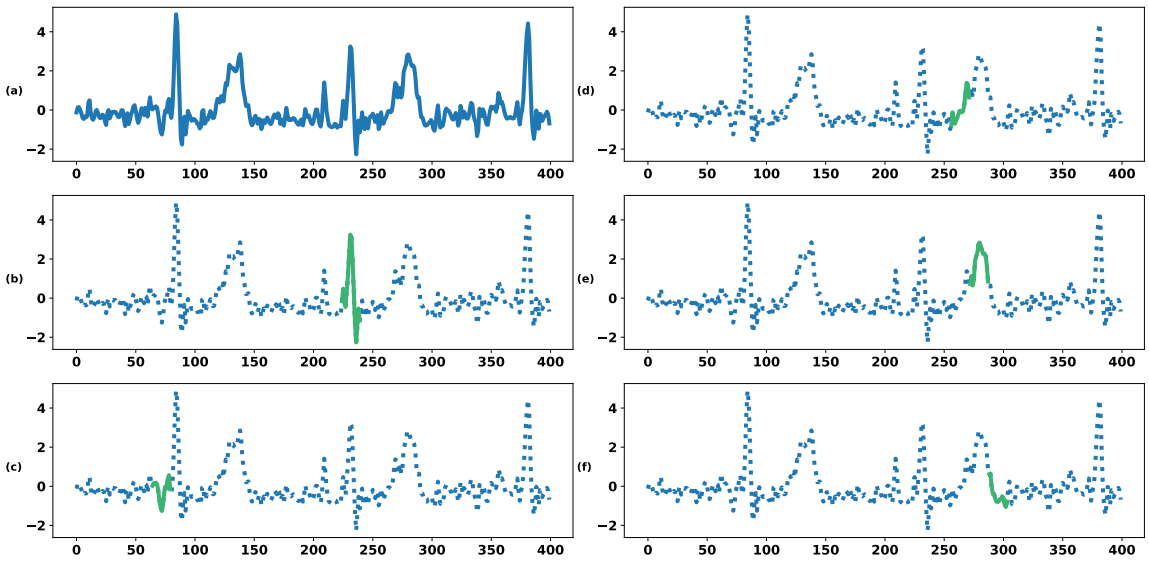

Figure 3.10: Attention map generated by the ESA module present in the final layer. The dotted part of the ECG plot has no attention and the solid part is given attention

3.3 Summary

This chapter presents a novel MSTDLNet model that learns the multi-scale morphological and temporal representation from ECG signal for person identification. Specifically, we designed an innovative architecture using CNN and LSTM networks for multi-scale temporal representation learning along with a multitude of attention blocks for enhanced representation learning. The proposed model leverages the fine-to-coarse flow of information within a stacked convolutional network to learn the multi-scale temporal representation. The model gives state-of-the-art performance for person identification.

Experimental results suggest that the multi-scale temporal representation learned by the MSTDLNet model gives robust performance for *off-the-person* ECG records. The multi-scale temporal representation learned by the model exhibits better permanence property leading to performance improvement in multi-session analysis. Various attention modules proposed in the work significantly improve the biometric representation learned through blind segmentation process.

4

An Attentive Spatio-Temporal Learning Based Network for Cardiovascular Disease Diagnosis

Contents

Cardiovascular diseases (CVDs) remain a leading cause of mortality globally [2]. Therefore, it is crucial to monitor heart health. The monitoring of heart health has become increasingly crucial, especially with the heightened risks of cardiac abnormalities and related fatalities due to the COVID-19 virus [176]. Electrocardiogram (ECG) signal stands as the most favored non-invasive method for the evaluation and continuous monitoring of cardiac conditions. Nevertheless, the shortage of expert cardiologists poses a significant obstacle in utilizing ECG signals effectively for continuous monitoring and diagnosis. In response to this challenge, several automated cardiovascular disease diagnosis systems have been developed, aiming to provide efficient and accurate diagnostic solutions [80, 83, 117, 120, 124, 177, 178].

The ECG signal is a cyclostationary signal that records the heart's electrical activity. In a standard clinical setup, a 12-lead ECG recording is done, where each lead views the heart from a different angle. Thus, the 12-lead ECG signal varies across different leads as well as the time axis. The variation of the ECG signal across different leads shows the spread of the electric pulse across the heart at a given moment. While the variation across the time axis represents the flow of the electric signal at a location. This spatio-temporal variation constitutes a three dimensional view of the heart's electrical activity [21]. Section 1.2 details the pathological manifestation of the CVD in multi-lead ECG signal, outlining distinct alterations in ECG waveforms and segments across specific leads corresponding to various CVDs. Along with the spatial variation of the ECG signal, the morphological shape of the ECG waveforms and ECG segments change according to specific CVD. Thus the variation of the clinical components of the ECG signal across different leads as well as along the temporal scale constitute the major cue for diagnostic decision making [21].

A detailed survey of the existing automated diagnostic models is provided in Section 1.4. The existing deep learning based methods are designed to learn the lead specific information in the first stage, and then the learned information is fused to obtain a diagnostic decision. Thus, the existing methods lack in fully exploiting the concurrent spatio-temporal variation present in the ECG signal. This has motivated us to design a neural network model that can learn the representation by leveraging the concurrent spatio-temporal variation of the ECG signal.

In this Chapter, we have proposed an attentive spatio-temporal learning based neural network (ASTLNet) to effectively learn the concurrent multi-scale spatio-temporal representation from the ECG signal. The proposed architecture consists of two modules, (a) Spatio-Temporal Representation

Learning (STRL) module, and (b) Attentive Spatio-Temporal Aggregation(ASTA) module. The STRL module consists of a clustered multi-head criss-cross attention (MHCCA) interleaved within a hierarchical LSTM (HLSTM) network. The clustered MHCCA layer facilitates learning the spatio-temporal representation by aggregating the local temporal representations. The ASTA module is introduced to effectively aggregate the multi-scale spatio-temporal representation learned by the STRL module. The STRL module consists of two recurrent MHCCA layers followed by a novel multi-aligned attention (MAA) layer. The MAA layer is introduced to obtain multiple context vectors by giving more weight to diagnostic significant regions in the temporal dimension. The proposed model is tailored for a multi-label CVD diagnosis application, enabling the diagnosis of multiple cardiovascular diseases within a single subject.

4.1 Preliminaries

In this section, we have formally defined the problem and introduced the architecture of the MHCCA layer for better comprehension of the proposed framework.

4.1.1 Problem Statement

In this work, the multi-lead ECG based CAD diagnosis problem is formulated as a multi-view time series classification problem. The ECG signal from each lead represents one view of the heart. The ECG data from n^{th} lead can be represented as, $x^n =< x_1^n, x_2^n, .., x_t^n, .., x_T^n >$, where $x_t^n \in \mathbb{R}^{d_i}$ is the input vector at time location t. d_i is the input vector dimension and $1 \leq n \leq N, N = 12$.

Given $X =< x^1, x^2, .., x^N >$, the CVD diagnosis task is to estimate the presence of a cardiac abnormality y_k in a sample space of Y, where $Y = [y_1, .., y_k, .., y_K]$. $y_k \in [0, 1]$, where 0 and 1 represent the disease's absence and presence, respectively.

4.1.2 Multi Head CC Attention (MHCCA)

The MHCCA network is introduced to aggregate the information across different leads as well as along the time-axis. The MHCCA network is designed by introducing multi-head attention to the CCNet [179]. Figure 4.1 gives a visual description of the criss-cross attention operation. As shown in Figure 4.1, a linear operation is done on the input feature map $I \in \mathbb{R}^{N \times T \times d}$ to obtain the three feature maps Q^h(Query) = $W_Q I$, K^h(Key) = $W_K I$, and V^h(Value) = $W_V I$ corresponding to each head h. Here,

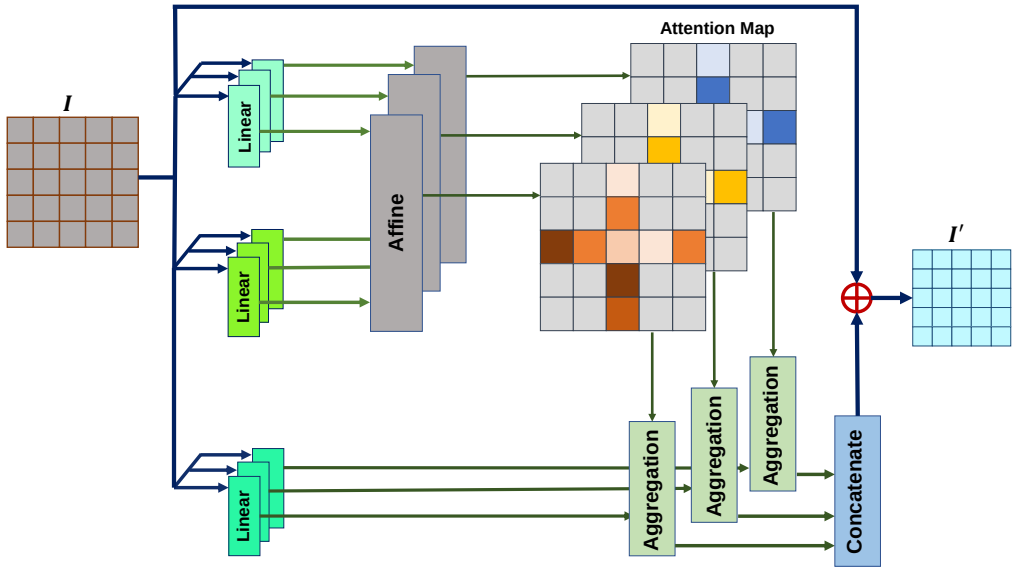

Figure 4.1: Flowchart of the multi-head CC attention (MHCCA)

$\{Q^h, K^h, V^h\} \in \mathbb{R}^{N\times T\times d_k}$ and $\{W_Q, W_K, W_V\} \in \mathbb{R}^{d\times d_k}$. For n^h number of heads $d_k = d/n^h$. Following this, the attention map $A^h \in \mathbb{R}^{(N+T-1)(N\times T)}$ is generated through an affinity operation. The mathematical operation to obtain an element $a_{i,u}$ of A^h is given in eq. 4.1.

$$a_{i,u} = softmax(Q_u^h \omega_{i,u}^T) \tag{4.1}$$

Here, $Q_u \in \mathbb{R}^{d_k}$ stands for the feature vector of Q^h corresponding to a spatial location u. The feature vector $\omega_{i,u}$ stands for the i^{th} feature vector of the feature map $\omega_u \in \mathbb{R}^{(N+T-1)\times d_k}$. The feature map ω_u is obtained by extracting all the feature vectors of K^h corresponding to the same view and temporal axis as the location u.

Another feature map $\theta_u \in \mathbb{R}^{(N+T-1)\times d_k}$ is obtained from V^h for aggregating the contextual information. The feature map θ_u is obtained by extracting all the feature vectors of V^h corresponding to the same view and temporal axis as the location u. Finally, the MHCCA based contextual information map is obtained by using eq. 4.2 and 4.3.

$$I_u^i = \sum_{j=0}^{N+T-1} A_{j,u}^h \theta_{j,u}^h + I_u \tag{4.2}$$

$$I_u^{'} = concat[I_u^1, I_u^2, .., I_u^{n^h}] \tag{4.3}$$

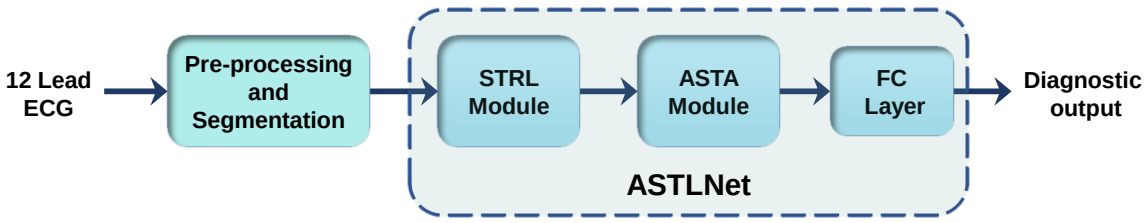

Figure 4.2: Flow diagram of the proposed CVD diagnosis method

4.2 ASTLNet for CVD Diagnosis

In this section, we have described our proposed framework. Figure 4.2 shows the flow diagram of the proposed methodology. First, the 12-lead ECG signal is filtered and normalised, followed by a blind segmentation process to extract the input vectors. These input vectors are fed to the proposed ASTLNet model to obtain the multi-scale spatio-temporal representation. Finally, the diagnostic report is obtained by passing the learned representation through a fully connected layer. Next, we have described different modules of the proposed framework along with their significance in the proposed framework.

4.2.1 Preprocessing and Segmentation

In the preprocessing stage, the ECG signal is first resampled to 200Hz. Subsequently, the ECG signal is passed through a band-pass butterworth filter with a lower cut-off frequency of 0.6 Hz and a higher cut-off frequency of 60 Hz to remove baseline wander and high frequency noises.

Following the filtering process, the ECG sequences are extracted from the filtered ECG signal, which is used to train and evaluate the proposed framework [171]. The ECG sequences are extracted using a rectangular window with an overlap of 0.75 fraction [171]. In the case of the PTB-XL database, the overlap fraction is kept at 0.25. An optimal length of the rectangular window is chosen for different datasets, i.e., $10s$, $2.5s$, and $30s$ for PTB, PTB-XL and CPSC-2018 datasets, respectively. Subsequently, the ECG sequences are further segmented into smaller ECG segments, followed by a z-score standardisation process [171]. The smaller ECG segments are extracted using a 100 ms rectangular window with an overlap of 0.75 fraction. The smaller ECG segments allow the LSTM model to effectively capture the temporal variation [180]. The process followed is similar to our work in [171]. The ECG segments of an ECG sequence are given as the input vectors to the LSTM cells

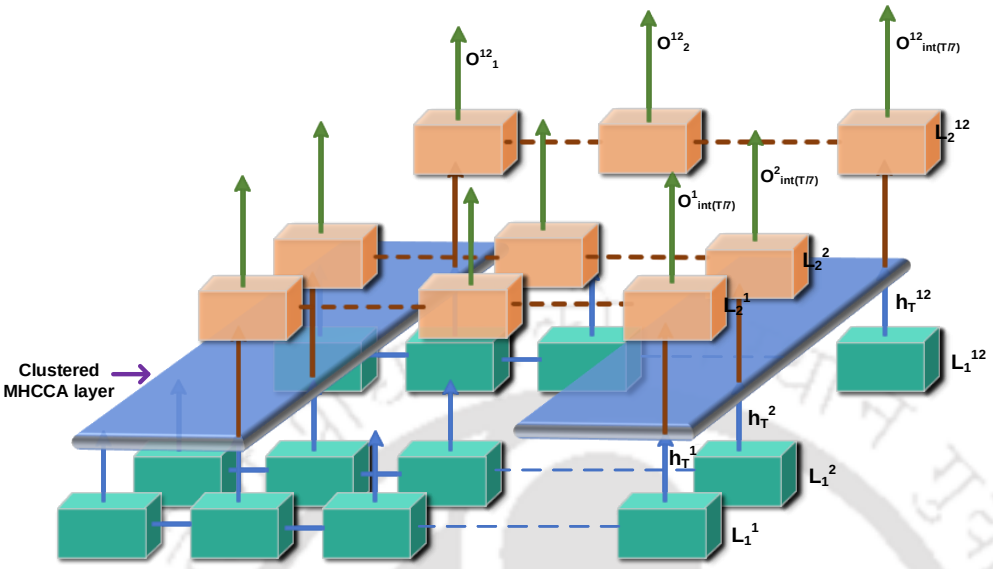

Figure 4.3: Architecture of the STRL module

of the proposed ASTLNet.

4.2.2 Proposed ASTLNet

4.2.2.1 Spatio-Temporal Representation Learning(STRL) Module

The STRL module is designed to learn the multi-scale spatio temporal information pertaining to the ECG signal. Figure 4.3 shows the architecture of the STRL module. The STRL module is composed of two layers of LSTM networks with a dilated connection. A clustered MHCCA layer is introduced between the two LSTM layers for aggregating the spatio-temporal information. The first layer (L_1) of the STRL module is composed of twelve LSTM networks $(L_1^1, L_1^2, .., L_1^n, .., L_1^{12})$ for learning the lead specific local morphological representations, i.e. P wave, QRS complex, ST-T segment etc. Eq 4.4 shows the output of the L1 layer corresponding to the n^{th} lead.

$$< h_1^n, h_2^n, .., h_t^n, .., h_T^n >= L_1^n(x_1^n, x_2^n, .., x_t^n, .., x_T^n) \tag{4.4}$$

where $h_t^n \in \mathbb{R}^{d_h}$ and d_h is the hidden dimension of L_1 layer. Following this, the outputs of L_1 are normalized, and clusters are formed by taking 7 consecutive local representations of all the 12 leads. Each cluster is passed through a MHCCA module followed by an average pooling and dropout function. The MHCCA module will aggregate the spatial and temporal information respective to each

local representation of a cluster. This operation is shown in eq. 4.5, and eq. 4.6;

$$\hat{h}_c = \{h^n_{7\times(c-1)+k} \mid n = \{1, .., 12\}, k = \{1, .., 7\}\} \tag{4.5}$$

$$\tilde{h}_c = AvgPool(MHCCA(\hat{h}_c)) \tag{4.6}$$

Here $c \in [1, int(T/7)]$. Finally, the learned spatio-temporal representations are passed to the second layer (L_2) of lead specific dilated LSTM networks ($L^1_2, .., L^n_2, .., L^{12}_2$) followed by normalization. This layer is introduced to learn the multi-scale spatio-temporal representations. This is shown in eq 4.7.

$$< O^n_1, .., O^n_c, .., O^n_{int(T/7)} >= L^n_2(\tilde{h}^n_1, .., \tilde{h}^n_c, .., \tilde{h}^n_{int(T/7)}) \tag{4.7}$$

where $O^n_c \in \mathbb{R}^{dO}$ and d_O is the hidden dimension of L_2 layer. This layer is introduced to learn the inter-morphological variation of the ECG signal with embedded spatial variation. Thus the proposed STRL module models the multi-scale spatio-temporal variation of the ECG signal for effective CAD diagnosis.

4.2.2.2 Attentive Spatio-Temporal Aggregation(ASTA) Module

The ASTA module is designed to aggregate the representations ($O^n_1, .., O^n_{int(T/7)}$), learned by the STRL module. As shown in Figure 4.4, the ASTL module is composed of two successive MHCCA layers followed by an MAA module. The successive MHCCA layers are introduced to effectively aggregate the multi-scale representation corresponding to each representation vector while accounting for the significant clinical information. Then, the outputs of the MHCCA layer are normalized and passed through an average pooling function to combine the multi-lead information respective to each time stamp, which is represented as $H = \{H_1, .., H_c, .., H_{int(T/7)}\}$, where $H_c \in \mathbb{R}^{dO}$. Finally, the aggregated outputs are given to the novel MAA module.

Multi Aligned Attention(MAA): This module extracts multiple diagnostic representations for the multi-label multi-class classification problem. To obtain the MAA weights, first, the representations encoded at each time stamp are linearly transformed to a latent space of dimension d_O. This is shown in eq 4.8;

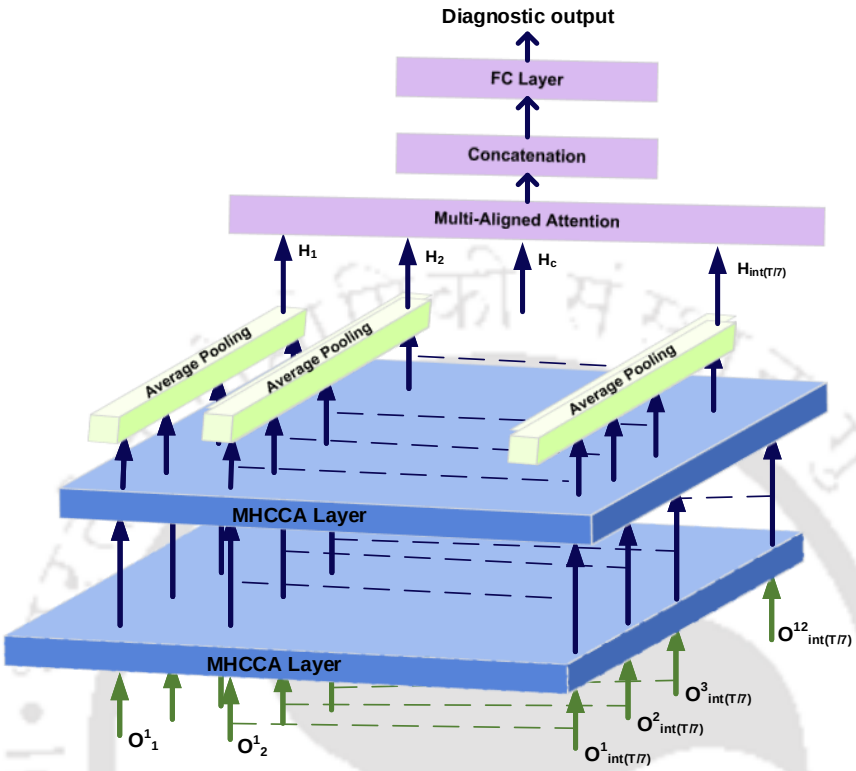

Figure 4.4: Architecture of the ASTA module

$$\hat{H} = W_l H + b_l \tag{4.8}$$

Here, $\hat{H}_c \in \mathbb{R}^{dO}$, and $W_l \in \mathbb{R}^{dO \times dO}$. The matrix $\hat{H} = [\hat{H}_1, .., \hat{H}_c, .., \hat{H}_{int(T/7)}]$ is formed from the latent vectors. Following this, the latent vectors $\hat{H}_c$ are used to obtain the correlation factor $r_c = \{r_{c,1}, .., r_{c,m}, .., r_{c,M}\}$ with the learnable dictionary $K = [K_1, .., K_M]$. The learnable dictionary $K \in \mathbb{R}^{dO \times M}$ has M keys. This can be expressed in eq 4.9;

$$r_{c,k} = \hat{H}K \tag{4.9}$$

Finally, the correlation factors $r_{c,k}$ are fed to a softmax activation function to obtain attention weights corresponding to a temporal location. This is given in eq 4.10. Finally, the diagnostic representation vectors $D_m \in \mathbb{R}^{dO}$ are obtained using eq 4.11.

$$\alpha_{c,m} = \frac{exp(r_{c,m})}{\sum_{c=1}^{C} exp(r_{c,m})} \tag{4.10}$$

$$D_m = \sum_{c=1}^{c=int(T/7)} \alpha_{c,m} H_c, \text{ for } 1 \leq m \leq M \tag{4.11}$$

We have introduced a regularisation function to enforce the dissimilarity in the alignments. The regularisation function is obtained by taking the cosine similarity between the multi-aligned attention weight arrays, $\alpha_m = [\alpha_{1,m}, .., \alpha_{c,m}, .., \alpha_{int(T/7),m}]$. This is expressed in eq 4.12. From eq 4.12, it can be observed that the alignment loss (L_a) will become more when any two weight arrays are correlated.

$$L_a = \beta \sum_{i=1}^{M-1} \sum_{j=l+1}^{M} \frac{\alpha_i \cdot \alpha_j}{\|\alpha_i\| \cdot \|\alpha_j\|} \tag{4.12}$$

Finally, the probability of the presence of a disease is estimated by concatenating the diagnostic representation vectors D_m and passing it through a fully connected layer followed by a sigmoid function. The sigmoid function, which is used to estimate the probability in the multi-label binary-class detection scenario (PTB-XL and CPSC-2018 ECG datasets), is replaced by a softmax function for the multi-class classification scenario (PTB dataset).

4.2.3 Optimisation Method of ASTLNet

The proposed ASTLNet is optimised in an end-to-end manner using a joint loss function. The joint loss function is defined as the summation of diagnostic loss (L_d) and alignment loss (L_a) as in eq 4.13. In the case of the multi-label binary-class detection scenario, we have used a binary cross-entropy loss function which is defined as in eq 4.14. In the case of the multi-class classification problem, we have used a cross-entropy loss function.

$$L = L_d + L_a \tag{4.13}$$

$$L_d = \frac{1}{N} \sum_{n=1}^{N} y_n log(p(n)) + (1 - y_n)\ log(1 - p(n)) \tag{4.14}$$

Here, N stands for the number of labels (i.e. diseases) present in the the detection task, and $p(n)$ is the probability estimated for n^{th} disease label.

4.3 Experiments

4.3.1 Database Description

The proposed method is evaluated using three publicly available ECG databases; i.e., PTB [156], PTB-XL [181] and CPSC-2018 database [182]. A brief description of the dataset is given below.

4.3.1.1 PTB Database

The PTB ECG database contains 549 twelve-lead recordings taken from 290 subjects. Out of the 549 recordings, 368 recordings taken from 148 subjects have been diagnosed with MI, and 80 recordings taken from 52 subjects as healthy. All the recordings are digitised with a sampling frequency of 1000 Hz. We have used all the MI and healthy recordings for the MI detection task. In the case of MI localisation, we have used recordings that are annotated as anterior myocardial infarction (AMI), anteroseptal myocardial infarction (ASMI), anterolateral myocardial infarction (ALMI), inferior myocardial infarction (IMI), inferolateral myocardial infarction (ILMI), or healthy. We have used 312 MI recordings and 80 healthy recordings for the MI localisation task. In our experiment, we have stratified the dataset into five folds such that each fold will have recordings from different subjects. This is an inter-person evaluation strategy as in [113, 183].

4.3.1.2 PTBXL Database

The PTB-XL database comprises 21837 twelve-lead ECG records taken from 18885 subjects. All the recordings are of a duration of 10 seconds and sampling frequency of 500 Hz. The PTB-XL database has 21430 recordings annotated with 5 diagnostic superclasses (i.e. NORM, CD, HYP, MI, STTC) and 23 diagnostic subclasses. Annotations on rhythm are available for 21046 recordings. Each record in the database is annotated with one or more diagnostic labels. In this work, we have used the recommended train-test split for the evaluation of the proposed framework.

4.3.1.3 CPSC-2018 Database

The CPSC-2018 challenge database is a publicly available database with 6877 twelve-lead ECG recordings collected from 11 hospitals. The duration of the recordings varies from 6s to 60s, and the sampling frequency is 500 Hz. Each recording in the database is labelled with a minimum of one diagnostic label. The database has nine diagnostic labels, which include normal rhythm, conduction

disturbances, atrial fibrillation, and ST-T changes. In our experiment, we stratified the dataset into ten folds while maintaining an equal distribution of each label.

4.3.2 Evaluation Method

In this work, we have used threshold free area under the curve (AUC), F1-score (F1), accuracy (Acc), hamming loss (Hamm), precision (Pre), and sensitivity (Sen) as the performance metric. In the case of the multi-label multi-class classification scenario (i.e. PTB-XL and CPSC-2018 database), we have obtained the macro-averaged scores that give equal weights to all the diagnostic labels.

4.3.3 Implementation Details

4.3.3.1 Network Parameters

The ECG segments obtained through the blind segmentation process are given as input to the proposed ASTLNet model. The input vector (x_t^n) has a dimension d_i of 20. The hidden dimension d_h of L_1 layer is set at 90 and d_O of L_2 layer at 150. We have taken 7 consecutive hidden outputs of all the twelve leads to form a cluster in the STRL module. The number of heads n^h is set at 6 in this work. We have used 3 keys(M) in the MAA module. The dropout rate is set at 0.25 for all the layers.

4.3.3.2 Training Setting

All of the deep learning models in this work are implemented using PyTorch framework. We have used the Adam optimiser to optimise the model parameters. An optimal set of training hyperparameters for each dataset are obtained by multiple trials on the validation dataset. We have used an initial learning rate (lr) of 0.0005 for PTB and PTB-XL datasets. An initial lr of 0.015 is used for the CPSC-2018 dataset. The learning rate is varied to $\frac{1}{5}^{th}$ of the initial lr by using a cosine learning rate scheduler. The batch size is set at 100 for the PTB and PTB-XL datasets and 32 for CPSC-2018 dataset. The proposed ASTLNet model is trained for 50 epochs in the case of the PTB and PTB-XL datasets and 100 epochs for the CPSC-2018 dataset.

4.3.4 Baseline Models for Comparison

The effectiveness of the proposed ASTLNet architecture is validated by comparing the model's performance with DL based state-of-the-art methods. We have evaluated the performance of the standard DL based architectures, i.e. VGG16 [175], ResNet50 [172], and LSTM [149] on PTB-XL and

CPSC-2018 ECG databases. We have also implemented some state-of-the-art automated diagnostic methods, i.e. ATICNN [119], DMSFNet [115], and HLSTM [184], for the performance comparison. The results for the baseline models are obtained by using the same experimental setup as outlined in this work. In the case of the PTB database, we have compared our performance with the existing state-of-the methods on the respective database. Since the literature methods employ a similar experimental setup, we have directly taken the performance results for the comparison.

4.3.5 Experiment on PTB database

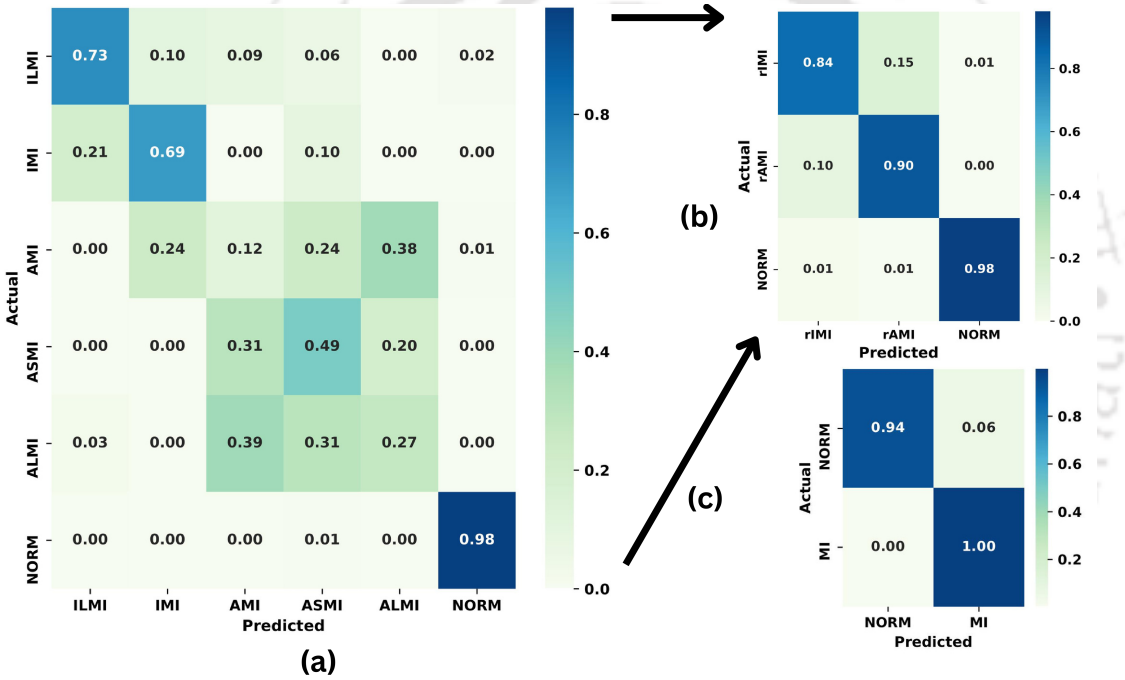

Figure 4.5: (a) Normalised confusion matrix for MI localisation task. (b) Normalised confusion matrix computed for relative locations, i.e. rAMI (includes ALMI, ASMI, and AMI), rIMI (includes IMI, and ILMI), and normal (c) Normalised confusion matrix for MI detection task.

We have obtained the average of the five-fold cross-validation results for the MI detection and localisation task using the PTB database. The performance of the proposed ASTLNet architecture is compared with the existing state-of-the-art works in Table 4.1. A few works have reported MI detection results using the challenging inter-person evaluation strategy. From Table 4.1, it can be observed that the proposed ASTLNet model gives a detection accuracy of 95.7%, which is significantly better than the existing works. In terms of F1-score, sensitivity, and specificity metrics, the proposed model performs better than the works that have used the complete dataset for evaluation. From Table 4.1, it can be observed that the proposed model gives superior performance for MI localisation tasks for all

the evaluation metrics. However, the classification performance degrades slightly in the case of the localisation task. This is mainly due to the misclassification within a relative location, i.e., AMI, ASMI, and ALMI of the anterior location (rAMI) and IMI and ILMI of inferior location (rIMI). For validation, the performance of the model is recalculated with the relative locations as the new labels, i.e. rAMI, rIMI, and Normal. We obtained an accuracy of 85.5%, precision of 84.81%, a sensitivity of 85%, specificity of 92.65%, and an F1 score of 83.61%. This suggests that the model gives robust performance both for the MI detection and localisation task.

The confusion matrix for the localisation and detection tasks are shown in Figure 4.5(a) and Figure 4.5(c). From Figure 4.5(a), it can be observed that the ASTLNet has very high precision in detecting healthy subjects. The confusion matrix for the new labels, which are obtained by grouping relative locations (i.e. rAMI, rIMI, and Normal) is given in Figure 4.5(b). The model is observed to perform well in detecting the relative locations of MI, i.e., inferior (rIMI) or anterior (rAMI). Figure 4.5(c) shows that the proposed model can accurately detect all the healthy cases, while a few MI classes are being misclassified as normal. This shows the proposed model gives a robust performance for the detection and localisation of MI diseases.

Table 4.1: Comparison With The Existing Works on MI Detection and Localization Using PTB Database

Works	Complete Dataset Used	Task	Acc	Pre	Sen	Sp	F1
Sharma and Sunkaria [185]	No	Detect	81.71	-	79.01	79.26	-
Reasat and Shahnaz [186]	No	Detect	84.54	-	85.33	84.09	-
Han and Shi [187]	No	Detect	92.69	-	80.96	86.14	-
Liu et. al. [120]	Yes	Detect	93.08	-	94.42	86.29	-
Han and Shi [113]	No	Detect	95.49	-	**94.85**	**97.37**	**96.92**
	Yes	Localise	55.74	-	47.58	55.37	47.94
Proposed	Yes	Detect	**95.7**	**94.1**	**94.7**	**94.7**	**92.1**
Proposed	Yes	Localise	**56.7**	**51**	**51.6**	**91.2**	**48.3**

4.3.6 Experiment on PTBXL database

The results obtained on the test fold of the PTB-XL database are tabulated in Table 4.2. The best performing model on the validation fold of the PTB-XL database, corresponding to each evaluation metric, is used to obtain the results on the test fold. The results of the proposed ASTLNet model are compared with the six baseline models across all the three diagnostic categories, i.e., superclass, subclass, and arrhythmia. From Table 4.2, it can be observed that the proposed model performs superior compared to the baseline models across all the three categories. The multi-scale CNN based model, DMSFNet, can be observed to perform slightly better in terms of the accuracy metric for the arrhythmia category. This may be because of the fact that the spatial variation preserves less diagnostic information for the arrhythmia labels present in the database. However, the proposed method gives superior performance in terms of the F1 score and the AUC metrics.

The performance of the HLSTM model, which encodes the multi-scale temporal information, and the vanilla LSTM model are given in Table 4.2. From Table 4.2, it can be observed that the HLSTM model performs better than the vanilla LSTM model. This shows that the HLSTM model encodes more diagnostic information by exploiting the multi-scale temporal variation. The proposed ASTLNet model gives superior performance compared to the HSLTM and LSTM models. This shows that the multi-scale spatio-temporal information learned by the ASTLNet model captures the diagnostic information better.

Table 4.2: Performance Comparison of The Proposed Method With The Baseline Methods on The PTB-XL Database

Works	Rhythm				Super diagnosis				Sub diagnosis			
	Acc	F1	Hamm	AUC	Acc	F1	Hamm	AUC	Acc	F1	Hamm	AUC
VGG16 [175]	89.16	32.88	1.62	90.82	62.14	70.72	11.87	90.93	53.31	37.77	3.65	88.67
Resnet50 [172]	88.06	45.10	1.70	89.77	60.75	70.43	11.96	91.01	51.96	42.82	3.64	92.41
ATICNN [119]	89.63	40.03	1.59	91.89	62.32	71.75	11.76	91.05	53.12	38.32	3.62	88.17
DMSFNet [115]	**90.01**	47.57	**1.45**	93.26	62.51	70.72	11.99	90.72	52.47	39.56	3.66	89.74
LSTM	87.78	41.95	1.83	92.93	60.15	70.65	12.06	90.49	48.59	40.83	3.72	89.22
HLSTM [184]	88.11	43.29	1.73	91.21	60.61	72.49	11.79	91.26	50.21	44.16	3.68	91.61
ASTLNet	**89.68**	**49.77**	**1.58**	**94.15**	**62.69**	**73.58**	**11.62**	**91.31**	**54.05**	**46.86**	**3.54**	**93.2**

Table 4.3: Performance of The Proposed Method Across Different Diagnostic Labels

	AFIB	AFLT	BIGU	PACE	PSVT	SARRH	SBRAD	SR	STACH	SVARR	SVTAC	TRIGU
AUC	98.6	88.3	95.1	94.5	99.9	85.3	95.9	92.7	99.4	86.2	99.3	94.7
GP	152	7	8	29	2	77	64	1678	82	14	3	2
TP	135	4	3	17	2	16	33	1648	73	2	1	2
PRE	87.7	80	25	94.4	33.3	21.1	78.6	91.8	85.9	7.7	9.1	0.1

	NORM	STTC	HYP	MI	CD
AUC	94.1	93.6	84.1	92.4	92.4
GP	964	523	263	553	498
TP	894	430	142	407	380
PRE	80.5	73.1	52.6	75.5	76.6

	NORM	ISC	ISCA	STTC	NST	LAO	RAO	LVH	RVH	SEHYP	PAMI	IMI	LMI	PMI	CLBBB	CRBBB	ILBBB	IRBBB	IVCD	AVB	LAFB	WPW
AUC	96.3	76.6	92.4	99.7	94.1	93.5	85.5	97.9	95.1	98.1	90.2	90.3	95.3	95.1	94.1	99.8	92.6	82.2	92.6	99.8	92.1	95.7
GP	964	128	40	223	77	42	10	214	12	3	308	328	20	2	54	54	8	112	79	83	180	8
TP	912	84	18	158	35	8	2	130	5	2	241	205	5	2	47	53	1	74	17	44	136	4
PRE	76.4	51.2	32.7	42.0	17.8	14.0	40.0	51.4	19.2	28.6	73.3	68.6	13.2	0.1	95.9	69.7	12.5	63.2	31.5	42.7	76.8	100.0

The performance of the model across different diagnostic labels is tabulated in Table 4.3. From Table 4.3, it can be observed that the model can accurately detect the sinus rhythm (SR) with a precision value of 91.8%. The model also has a very high recall value, i.e. 1648 records have been correctly detected out of 1678 records. The model performs well in detecting other sinus rhythms also, i.e., sinus bradycardia (SBRAD) and sinus tachycardia (STACH). It shows an AUC value of 99.4% and 95.9%, respectively, for SBRAD and STACH. From Table 4.3, the ASTLNet model can be observed to detect paced rhythms precisely. The ASTLNet model shows superior performance for life-threatening arrhythmias like atrial fibrillation (AFIB) and atrial flutter (AFLT) with a precision of 87.7% and 80%, respectively. The proposed model has an impressive AUC of 98.6% for AFIB. In the case of other supra-ventricular arrhythmias (SVARR), i.e. supra-ventricular tachycardia (SVTAC), paroxysmal ventricular tachycardia (PSVT), and SVARR, the performance of the model reduces slightly. This may be because of less amount of training data. The same gets reflected in the case of the model's performance for bigeminal (BIGU) and trigeminal (TRIGU) rhythms. However, the model shows an impressive AUC (99.9%) and recall(100%) score in detecting clinically important PSVT arrhythmia. Similarly, the model has an impressive AUC of 94.1% and precision of 80.5% for detecting the normal class in the diagnostic superclass category. The model has a high recall for normal class detection, i.e. 894 records out of 964 records are correctly detected. In the case of hypertrophy (HYP), the model gives a relatively lower performance. This may be because of the fact that hypertrophy can be diagnosed better using an echocardiogram signal than the ECG signal. Similar performance of the model can be observed in detecting subcategories of HYP, i.e., left atrial overload (LAO), right atrial overload (RAO), left ventricular hypertrophy (LVH), and right ventricular hypertrophy (RVH). In the case of the diagnostic subclass category, it can be observed that the model gives superior performance in detecting significant diagnostic labels, i.e. Normal (NORM), ST-T changes (STTC), ischemic ST-T changes (ISCA, ISCI), MIs (i.e. AMI, IMI), conduction disturbances (CD) (CRBBB, CLBBB). It can be observed that the model gives a good precision of 73.3% and 68.6% for AMI and IMI, respectively. Similarly, the model gives a robust performance of 95.9% and 69.7% in terms of precision value for CLBBB and CRBBB, respectively. However, the model performance is slightly off in the case of rare MIs, e.g. lateral MI (LMI) and posterior MI (PMI). Though PMI is difficult to diagnose using a 12-lead ECG recording system, the model gives a 100% recall score. The proposed ASTLNet model can be observed to perform well in the case of other CDs, i.e., left atrial fascicular

Table 4.4: Performance Comparison of The Proposed Method With The Baseline Methods on The CPSC-2018 Database

Works	Acc	F1	Pre	Sen	Ham	AUC
VGG16 [175]	71.5	68.5	76.9	65.2	5.0	93.7
Resnet50 [172]	71.4	73.0	77.2	70.9	4.5	94.8
ATICNN [119]	79.4	77.7	79.8	76.2	3.8	95.3
DMSFNet [115]	79.9	80.3	**83.3**	78.7	3.5	96.7
xECGNet [127]	68.9	77.4	-	-	-	-
LSTM	43.2	56.9	55.3	61.7	7.7	87.7
HLSTM [184]	75.6	78.9	80.3	78.5	3.7	96.7
ASTLNet	**80.0**	**81.8**	**81.0**	**82.9**	**3.5**	**97.0**

block (LAFB) and wolff-parkinson-white (WPW) syndrome.

4.3.7 Experiment on CPSC-2018 database

The results obtained by the baseline models and the proposed ASTLNet model on the CPSC-2018 dataset are tabulated in Table 4.4. The model is trained using eight folds and validated on one fold. Finally, the best performing model on the validation fold is used to evaluate the performance on the test fold. Performance on each fold is obtained by taking each one as a test fold. The average performance on the ten-folds is given in Table 4.4. From table 4.4, it can be observed that the proposed model gives superior performance compared to the baseline models. The ASTLNet model gives a very high average AUC of 97% on this large dataset. The model gives an average accuracy and an F1 score of 80% and 81.8%. From Table 4.4, it can be observed that the HLSTM model gives significantly better results than the vanilla LSTM model. This may be because the HLSTM model is designed to learn the multi-scale temporal information present in the ECG data. Table 4.4 shows that the proposed ASTLNet model, which is designed to learn the multi-scale spatio-temporal representation, gives superior performance compared to the HLSTM model. This validates the effectiveness of the proposed framework.

The average value of the performance metrics across different diagnostic classes is shown in Table 4.5. It can be observed that the model gives robust performance for the critical diagnostic labels, i.e. atrial fibrillation (AFIB), bundle branch blocks (CLBBB and CRBBB), ST depression (STD). The model gives an AUC and an F1-score of 98.47% and 92.32%, respectively, for atrial fibrillation (AFIB). The model also gives a good performance in detecting healthy records (NORM). The model shows

Table 4.5: Average Performance of The Proposed Method Across Different Diagnostic Labels

	NORM	AFIB	1AVB	CLBBB	CRBBB	PAC	VPC	STD	STE
AUC	96.84	98.47	98.46	98.40	98.97	95.94	97.71	96.77	91.83
SEN	79.94	93.10	87.67	91.56	95.64	77.59	83.07	82.25	55.45
PRE	78.37	91.67	85.76	87.40	91.90	74.50	84.77	77.68	57.26
F1	79.00	92.32	86.66	89.21	93.71	75.69	83.73	79.81	55.79
ACC	94.00	97.13	97.09	99.18	96.40	95.59	96.93	95.06	97.38

an AUC score of 96.84% and an F1 score of 79% for detecting the NORM label.

4.3.8 Effect of Clustered CC Attention

The clustered CC attention is introduced in the model to effectively learn the spatio-temporal variation present in the ECG signal. To show the effectiveness of the proposed method, we have compared the performance of the proposed ASTLNet model with Model1. Model1 is implemented by removing the clustered CC attention layer from the STRL module of the ASTLNet model. The performance comparison is shown through a bar plot in Figure 4.6. Figure 4.6 shows the performance comparison for the three databases in terms of the F1-score and accuracy. The results on the PTB database are shown for the MI localisation task. From Figure 4.6, it can be observed that the introduction of the clustered CC attention significantly improves the performance of the model. The accuracy can be observed to improve by over 4.37% and 6.33% for the CPSC database and the PTB database, respectively. Similarly, the F1-score improves by 2.92% and 6% for the CPSC database and the PTB database, respectively. This shows that the spatio-temporal information of the ECG signal can be effectively captured by introducing the clustered CC attention layer. The spatio-temporal information learned by the clustered CC attention layer can significantly improve diagnostic performance.

4.3.9 Effect of Multi-Head CC Attention

In this work, we have introduced multi-head attention to the criss-cross attention mechanism. The multi-head attention mechanism is introduced to learn the spatio-temporal information by giving more weight to multiple diagnostic cues present in both the dimensions. The effectiveness of the multi-head CC attention is validated by evaluating the model performance for three cases, i.e., number of head n^h = 1, n^h = 3, n^h = 6. The performance is evaluated on the PTB and PTB-XL databases, and the results are shown in the Figure 4.7(a) and 4.7(b). From Figure 4.7 (a) and 4.7(b), it can be observed

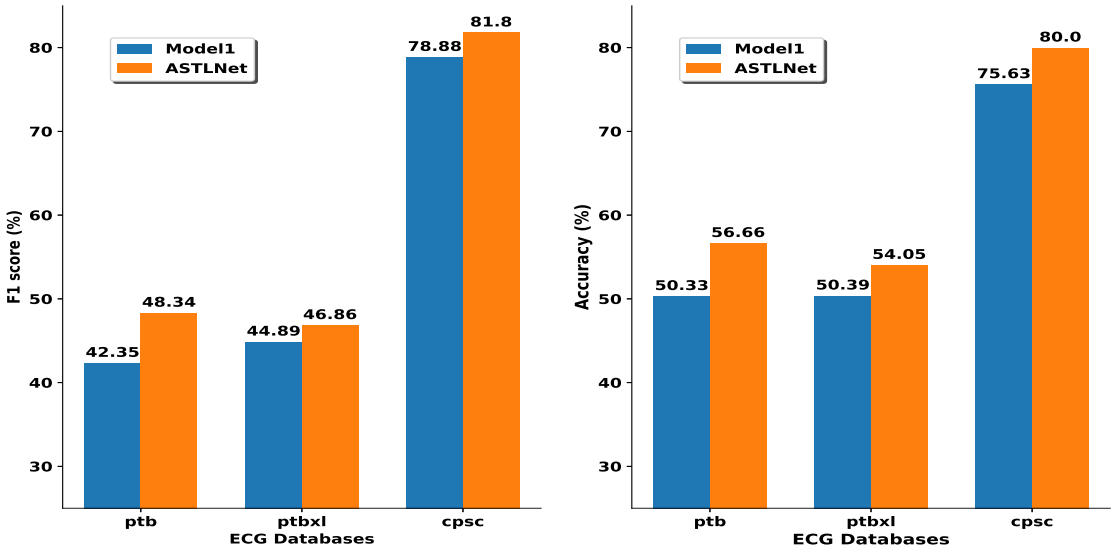

Figure 4.6: Comparison of the performance of the ASTLNet model without clustered CC attention (Model1)

that the model performs best for n^h = 6. It can also be observed that the model performs better with $n^h = 6$, compared to $n^h = 1$. This shows that the multi-scale spatio-temporal learning is better with multi-head CC attention.

4.3.10 Effect of Multi-Aligned Attention

In this work, we have proposed a novel multi-aligned attention mechanism. The module is introduced to learn multiple diagnostic representation vectors corresponding to a given ECG sample. Multiple representation vectors can encode manifold diagnostic cues present along the temporal dimension. To analyse the effectiveness of the MAA module, we have obtained the performance of the ASTLNet model for a different number of keys, i.e., $M = 1$, $M = 3$, and $M = 5$. The results obtained on the PTB and PTB-XL databases have been shown in Figure 4.7(c) and 4.7(d). From Figure 4.7(c) and 4.7(d), it can be observed that the model performs best for $M = 3$. This shows that multiple representative vectors can encode the diagnostic information better.

For better visualisation of the MAA module, we have shown the heat map of the attention weights in Figure 4.8. Figure 4.8(a1) and (a2) show the ECG signal of two different subjects taken from the PTB-XL database. In the case of Subject-1, it can be observed that a wide S-wave and a tall R wave are introduced in the ECG signal due to conduction disturbances (i.e. AVB, IRBBB) and hypertrophy (i.e. RVH). From the heat map, it can be observed that the first attention map (Figure 4.8(b1)) gives

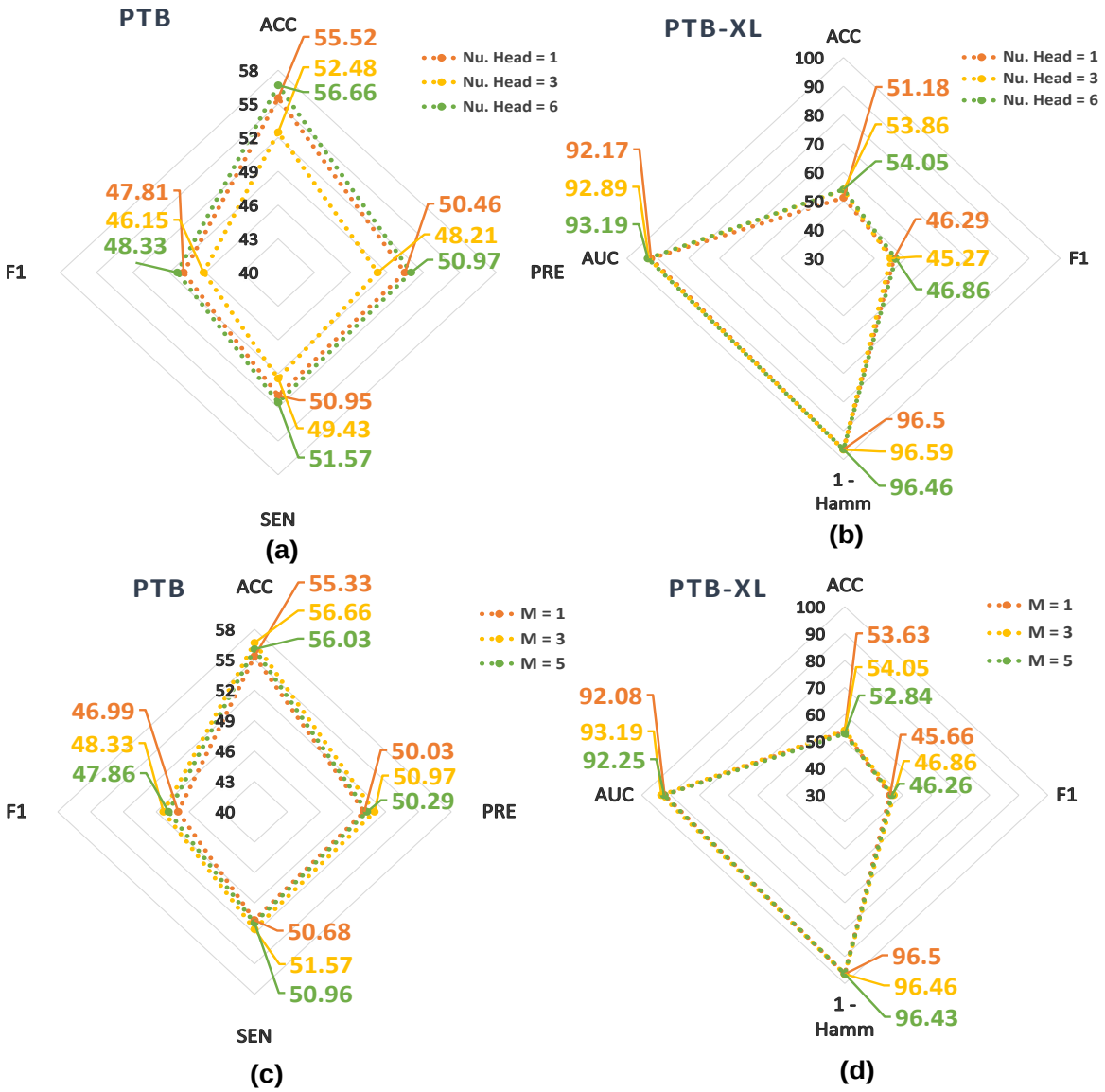

Figure 4.7: Comparison of performance of the ASTLNet model for different number of keys (M) and number of heads (n^h: Nu. Head)

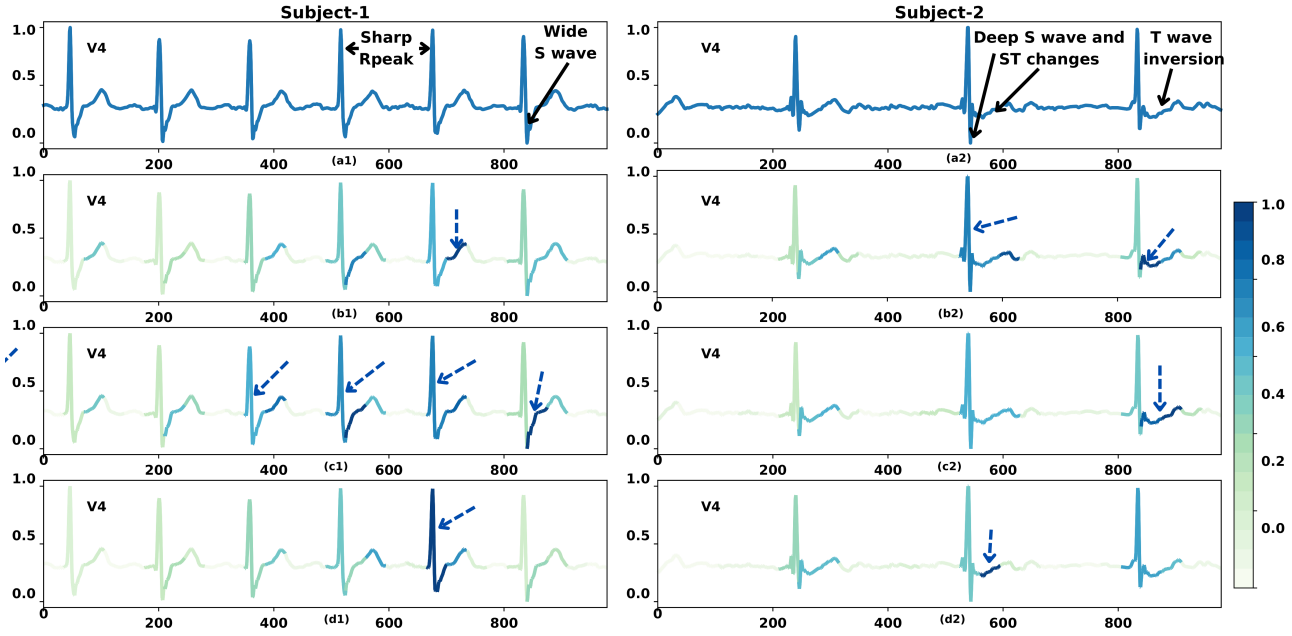

Figure 4.8: ECG records taken from two different subjects and their corresponding heat maps of the attention weights generated by the MAA module (number of keys $M = 3$).

more weight to the T wave, and the second attention map (Figure 4.8(c1)) gives more weight to the QRS complex and, in particular, the S wave. While the third attention map (Figure 4.8(d1)) focuses on the QRS complex. Similarly, deep S-waves, non-specific ST changes, and T wave inversions have been introduced in the ECG record of Subject-2 (Figure 4.8(a2)). From Figure 4.8(b2), it can be observed that more weight is given to the QRS complex and, in particular, the S wave. In the second attention map (Figure 4.8(c2)), more weight is given to the ST-T segment and the T wave. While the third attention map (Figure 4.8(d2)) focuses on the ST-T segment. This shows that the MAA module generates an attention map with different alignments that can appropriately give more weight to the manifold diagnostic cues.

4.3.11 Effect of Spatio-Temporal Learning

The ASTLNet model is designed to learn spatio-temporal representation for better diagnosis of CVDs. In this subsection, we have validated the effectiveness of spatio-temporal learning on the diagnosis of various CVDs. We have compared the performance of the proposed ASTLNet model for different diagnostic labels. The performance comparison is shown in Table 4.6 and Table 4.7. Table 4.6 shows the F1-score obtained by the methods for five broad diagnostic labels present in the PTB-XL

database. From Table 4.6, it can be observed that the proposed model shows performance improvement across all the diagnostic labels. Although the performance improvement corresponding to MI and CD diseases is moderate, the proposed model can be observed to perform significantly better for ST-T changes (STTC) and hypertrophy (HYP). From Table 4.6, it can be observed that the ASTLNet model performs better than the HLSTM model. The HLSTM model is implemented by removing the MHCCA based clustered CC attention and ASTA module. This shows that the MHCCA based modules help in learning better diagnostic representation. The performance comparison, in terms of F1-score, for the CPSC-2018 database is shown in Table 4.7. From 4.7, it can be observed that the proposed ASTLNet model gives significant performance improvement for CDs, i.e. CLBBB and CRBBB, and ST changes, i.e., STD and STE. The proposed ASTLNet model performs better than the existing methods across all the diagnostic labels except 1AVB. This shows that the proposed model can diagnose CVDs better by exploiting the spatio-temporal variation of multi-lead ECG signals.

Table 4.6: Comparison of The Proposed Method for Different CVDs in PTB-XL Database

Works	NORM	STTC	HYP	MI	CD
Resnet50 [172]	84.83	74.53	45.79	72.85	74.18
ATICNN [119]	84.39	74.98	45.58	71.27	75.05
LSTM [149]	85.66	75.86	44.82	71.86	75.08
HLSTM [184]	85.47	75.65	50.91	74.02	76.39
ASTLNet	**86.21**	**77.41**	**53.28**	**74.54**	**76.46**

Table 4.7: Comparison of The Proposed Method for Different CVDs in CPSC-2018 Database

Works	NORM	AFIB	1AVB	CLBBB	CRBBB	PAC	VPC	STD	STE
Resnet50 [172]	72.72	91.22	85.68	85.91	92.63	29.32	77.52	75.27	46.93
ATICNN [119]	75.65	92.07	87.10	84.53	93.11	66.08	83.02	75.93	41.73
LSTM [149]	57.84	65.43	51.74	75.43	88.23	25.26	61.76	56.43	30.26
HLSTM [184]	78.70	91.43	**87.25**	86.67	92.39	71.45	79.08	76.88	45.99
ASTLNet	**79.00**	**92.32**	86.66	**89.21**	**93.71**	**75.69**	**83.73**	**79.81**	**55.79**

4.3.12 Model Parameters

The number of trainable parameters used to implement the proposed model, and the baseline models are given in Table 4.8. From Table 4.8, it can be observed that the ASTLNet architecture requires the least amount of trainable parameters, i.e., approximately 2.42 million parameters. The

VGG16 [175] and Resnet50 [172] use over six times more parameters than the proposed ASTLNet model. DMSFNet, which is based on multi-scale CNN architecture, has over 7 million trainable parameters. This shows that the proposed model is lightweight compared to the baseline models. The fewer parameters used in the model may help in the better generalizability of the model.

Table 4.8: Comparison of Number of Model Parameters

Models	VGG16 [175]	Resnet50 [172]	ATICNN [119]	DMSFNet [115]	ASTLNet
Parameters	25983383	16032791	4998798	7082823	2419542

4.4 Summary

This chapter delves into the exploration of learning spatio-temporal representation from the ECG signal for multi-label CVD diagnosis. We introduced the innovative ASTLNet model, which focuses on capturing the spatio-temporal variation in a multi-lead ECG signal. Notably, we designed an STRL module by integrating an MHCCA layer within the HLSTM architecture, and subsequently aggregated the learned spatio-temporal representations using the ASTA module. Additionally, the novel MAA module was introduced to emphasize multiple diagnostic vectors, each encapsulating different diagnostic cues from a given ECG sample.

The experimental results show the effectiveness of the spatio-temporal representations learned by the ASTLNet model. It enables more accurate diagnosis of CVDs while utilizing fewer model parameters compared to current state-of-the-art methods. This underscores the significance of learning spatio-temporal representation in the realm of automated CVD diagnosis.

5

Identity ECG (iECG) Based Feature Conditioning for Person Adaptive CVD Diagnosis

Contents

The growing global prevalence of cardiovascular diseases (CVDs) and associated mortality is a primary concern for healthcare professionals [188]. Mitigating fatalities caused by CVDs and improving patient outcomes necessitates early and precise detection of cardiac abnormalities to facilitate timely interventions. Electrocardiogram (ECG) is a primary clinical non-invasive method used for diagnosing majority of cardiac abnormalities. The diagnostic process requires an expert cardiologist to carefully interpret the pathological characteristics in the ECG recordings of a subject. However, this manual approach is arduous, time-intensive, and susceptible to human error, limiting its scalability for broader cardiac health assessment and continuous monitoring. This has motivated researchers to develop automated CVD diagnosis system to expedite informed clinical decision making process [81, 83, 99].

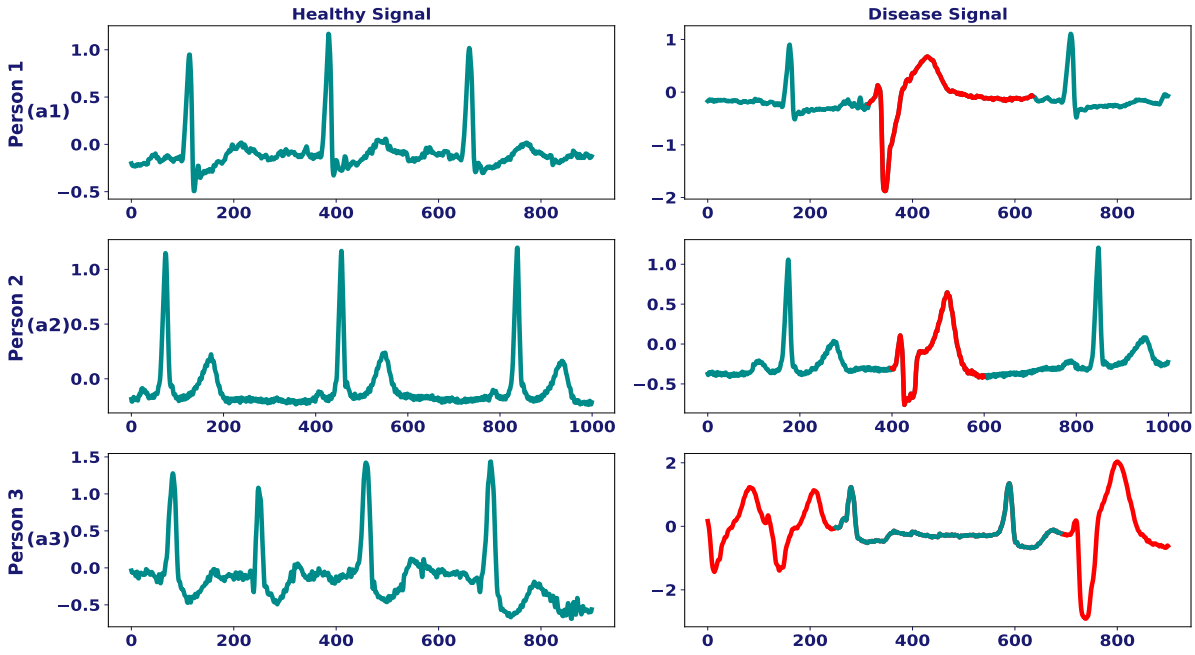

Figure 5.1: (a) Healthy ECG signal of three different subjects (b) PVC beats in the same subjects (marked in red colour)

Developing an automated CVD diagnosis system is a challenging task due to the variation in the morphological characteristics of the ECG signal across different subjects, termed as inter-individual variability [3]. The inter-individual variability is caused due to various generative factors, primarily due to physiological and geometrical differences in individuals' heart [3]. Figure 5.1 shows the healthy and premature ventricular Contraction (PVC) ECG beats of three subjects. The morphological characteristics of the healthy ECG signals of three subjects can be observed to exhibit significant variations. Similarly, the pathological characteristics of PVC ECG beats also exhibit substantial dif-

ferences across individuals, such as a deeper S wave for person 1 and a sharp T wave for person 2. The same underlying generative factors responsible for variation in the healthy ECG signal contribute to the variations in the pathological characteristics of diseased ECG signals. This variation leads to degradation in the diagnostic performance of an automated CVD diagnosis system [3, 7, 113]. This is because the trained global models are unable to generalize to data from a new test subject, generated by a different underlying distribution compared to the training data.

Several automated person-adaptive models for CVD diagnosis have been proposed in the literature, exploring a range of strategies to address the inter-person variation in the ECG signal. Section 1.4 of the Introduction chapter provides a comprehensive survey of existing person-adaptive diagnostic frameworks. However, these methods face several practical limitations. The existing methods necessitate both normal and diseased ECG data for each test subject to adapt a gloabl epert (GE) model and instantiate the local expert (LE) model. In a real-world health monitoring application scenario, diseased ECG data of a subject without any past cardiac disorder won't be available. Additionally, disease data for all potential cardiac conditions may not be available for a subject, potentially limiting the diagnostic model's performance. Another practical constraint in utilizing existing methodologies is the limited storage capacity, given the necessity to store person-specific models for each subject. Furthermore, the existing person-adaptive CVD diagnostic frameworks are designed for single-lead ECG signals, whereas multi-lead ECG signals are often used in hospital environments for diagnosing cardiac diseases.

In this chapter, we introduce an unsupervised adaptation framework utilizing the identity ECG (iECG) vector—a memory vector containing person-specific information—for the diagnosis of CVDs. Specifically, we introduced a conditioning network for affine transformation of diagnostic model's features conditioned on patient specific information. The patient specific information are encapsulated in a knowledge space through a memory module that utilizes ECG-based biometric representation (iECG vector). For each new subject a memory vector is generated by an auxiliary attention network that leverages the person-specific information encapsulated in the knowledge space. The auxiliary attention network is trained along with the main network and the conditioning network. We have introduced an auxiliary loss to enhance the learning of a more effective person-specific representation by the memory module. The proposed framework offers a modular design, seamlessly integrated to adapt different CVD diagnosis networks.

5.1 Proposed Framework

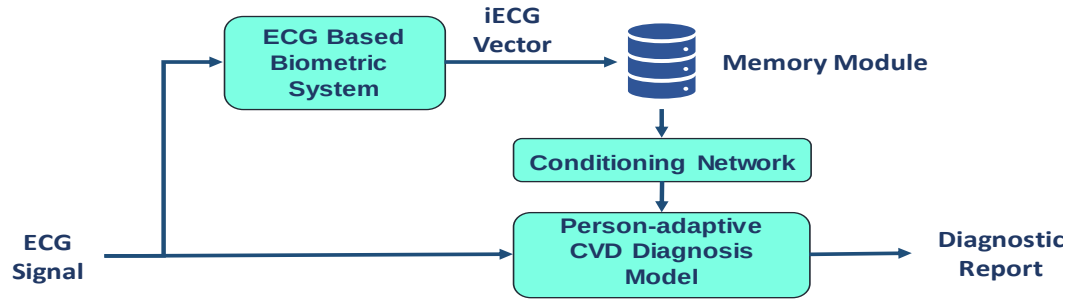

Figure 5.2: Basic block diagram of the proposed person-adaptive CVD diagnosis system.

The proposed person-adaptive CVD diagnosis framework comprises three fundamental modules: (i) Main Diagnostic Network, (ii) Memory Module, and (iii) Conditioning Network. The basic block diagram of this framework is illustrated in Figure 5.2. The person-adaptive training process involves storing iECG memory vectors, generated by an ECG-based biometric system, in a memory bank. Subsequently, for each test subject, an aggregated iECG (Agg_iECG) is generated by leveraging the stored memory vectors. This methodology offers the advantage of enabling the unsupervised adaptation of diagnostic features to new, unseen test subjects. Subsequently, the person-specific information is infused into the main diagnostic network through the conditional normalization of diagnostic features, facilitated by the conditioning network. The detailed architecture and functioning of the different modules are presented in subsequent subsections.

5.1.1 Main Diagnostic Network

The main diagnostic network is tasked with learning the diagnostic representation from the ECG signal. In the proposed framework, the main diagnostic network could be any variant of CNN or RNN based networks proposed in the existing literature. The diagnostic network typically comprises multiple layers of neural connections, designed to extract diagnostic features across varying levels of abstraction. Figure 5.3 illustrates the architecture of the proposed person-adaptive diagnostic framework, including the main diagnostic network, memory module, and conditioning network. Figure 5.3 depicts the main diagnostic network, featuring N layers of deep connections and a fully connected (FC) layer. In this work, we incorporated five main diagnostic networks: Vanilla LSTM [149], HLSTM [184], ResNet-18 [114, 172], ResNeXt-50 [174], and ASTLNet [189]. Below, we provide a

brief description of these networks tailored to their relevance in this study.

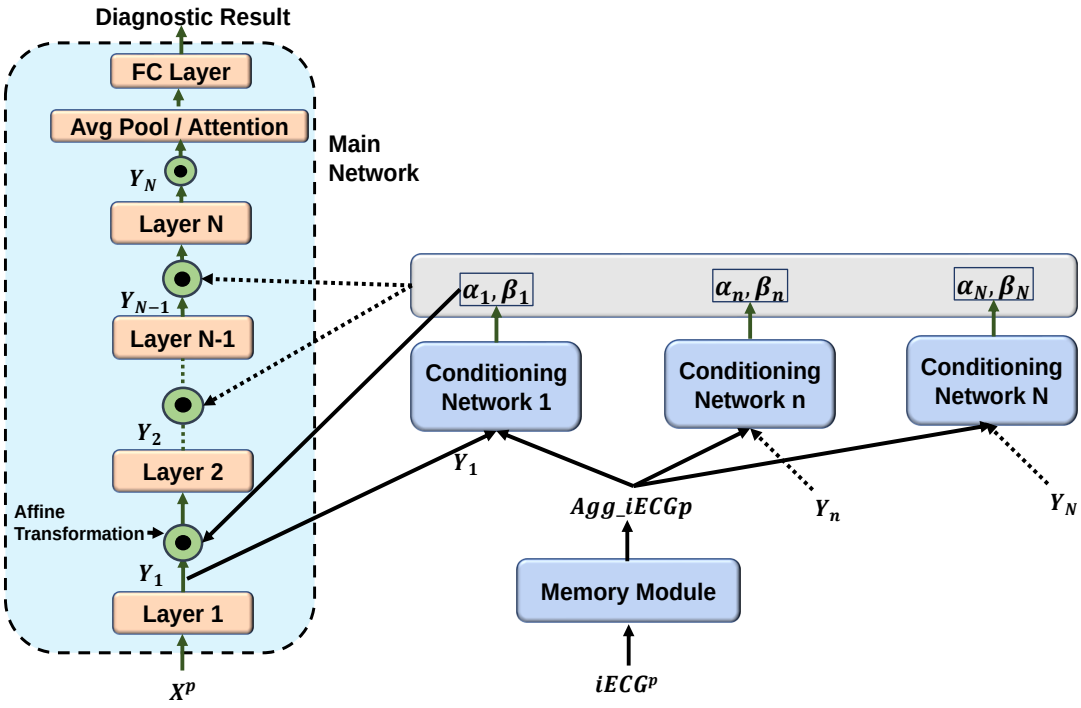

Figure 5.3: Architecture of the proposed person-adaptive diagnostic framework, including the main diagnostic network, memory module, and conditioning network.

HLSTM: The HLSTM model is composed of two layers of LSTM connected with different update intervals, as detailed in [184]. Within this architectural framework, conditional normalization is applied to all hidden vectors from both the first and second layers of LSTM following batch normalization.

Vanilla LSTM: The Vanilla LSTM model features a single layer of LSTM network. The diagnostic framework is similar to our work in [184, 189], where ECG segments of 0.1 sec duration are sequentially fed to the network at each time-stamp. Here, the final hidden vector of the LSTM network undergoes conditional normalization before being provided as input to the FC layer.

ResNet-18: The ResNet-18 architecture comprises four layers of residual blocks [114, 172]. The output feature map of each block undergoes conditional normalization before being passed as input to the subsequent layer.

ResNeXt-50: The ResNeXt-50 architecture comprises four layers of residual blocks [174]. The conditional normalization is applied to the output feature map of each block.

ASTLNet: In the ASTLNet model, we implement conditional normalization on the hidden outputs from two layers of the LSTM network within the spatio-temporal representation learning (STRL) module [189]. Conditional normalization is employed subsequent to the batch normalization of the hidden

outputs.

In this work, we propose the conditional normalization of features at each layer of the diagnostic network to facilitate patient-aware diagnostic feature learning. We posit that efficient modulation of intermediate features, conditioned on person-specific information will enhance the diagnostic network's ability to learn meaningful diagnostic features tailored to each individual [190]. A parallel can be drawn between the conditional normalization technique and other methods such as the Squeeze and Excitation network [170], Hypernetworks [191], and Convolutional sequence-to-sequence machine translation [192]. The conditional normalization selectively modulates the learned features based on the person-specific information. The input to the network is a blindly segmented ECG signal of a fixed duration, represented as X^p, where $X^p = \langle x(1), x(2), .., x(t), .., x(T) \rangle$ [171, 189]. Here, T denotes the number of sample points in the input ECG segment. The feature map output of layer n is denoted as Y_n, where $n \in 1, 2, ..., N$ and $Y_n \in \mathbb{R}^{d_n}$. Here, N represents the number of layers in the main diagnostic network, and d_n is the dimension of the feature map output of layer n. The conditional normalization of feature map, Y_n, is carried out by applying affine transformation on it. The mathematical expression of the affine transformation is expressed in eq 5.1.

$$Y_n^{'}(t) = Y_n(t) \odot \alpha_n(t) + \beta_n(t) \tag{5.1}$$

The affine transformation involves feature-wise multiplication and summation of parameters estimated by the conditioning network, denoted as $\alpha_n(t)$ and $\beta_n(t)$, respectively, with feature vector $Y_n(t)$, at time-stamp t. Here, $t \in 1, 2, ..T_n$ and $\alpha_n(t), \beta_n(t), Y_n(t) \in \mathbb{R}^{d_n}$.

5.1.2 Memory Module

The memory module is a fundamental building block of the proposed framework, encapsulating the knowledge space of person-specific information. The diagnostic features of the main network undergo affine transformation, conditioned on the person-specific information derived from the memory module. The proposed memory module consists of two key components: memory bank and an auxiliary attention network. The memory bank stores the iECG vectors, which represent person-specific information. The iECG vectors are generated using a pre-trained ECG-based biometric identification system. In this study, we utilized the HLSTM-based biometric system [184] as the pre-trained model. The biometric identification system is pre-trained with the healthy ECG signals from the PTB

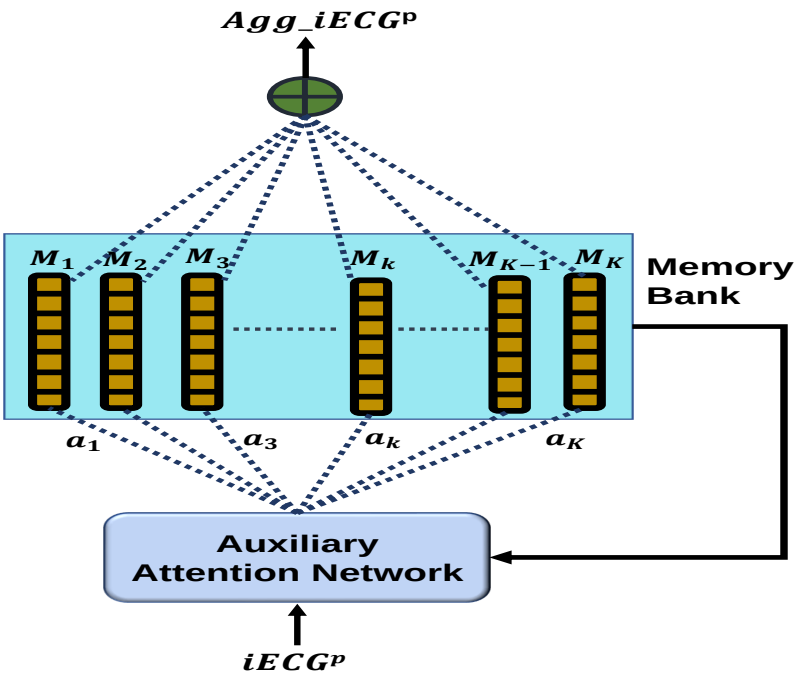

Figure 5.4: Architecture of the Memory Module

dataset and CPSC dataset (recordings exceeding 26s are considered). A dedicated biometric system is trained for each of the 12-lead ECG signals. Subsequently, iECG vectors are extracted for the same subjects from PTB [156] and CPSC datasets [182] using the pre-trained model. Additionally, iECG vectors from MIT-BIH arrhythmia [193] and Staff datasets [182] are also extracted. To form the memory bank, a K-means clustering algorithm is applied to cluster iECG vectors, and the resulting cluster centers (K) serve as the final memory vectors (M_k). This procedure is repeated for each lead, establishing dedicated memory banks. This approach allows the memory bank to represent the knowledge space pertaining to person specific information, facilitating the representation of person specific information for subjects not covered in the training dataset.

The key to unsupervised adaptation lies in deriving an aggregated iECG ($Agg\_iECG \in \mathbb{R}^{d_M}$), representative of a new test subject using the memory vectors, $M_k \in \mathbb{R}^{d_M}$, stored in the memory bank. Here, d_M represents the dimension of memory vectors. We hypothesize that a large pool of iECG vectors can represent the knowledge space encapsulating the person-specific information and a weighted sum of these stored memory vectors aptly represent the person-specific information of a new subject. This framework finds inspiration in the principles of speaker adaptive training, as outlined in [194, 195]. The weights (a_k) are determined by the auxiliary attention network, and the

mathematical expressions governing the estimation of a_k are detailed in eq 5.2 and 5.3.

$$e_k^p = V^T \tanh(W^T iECG^p + U^T M_k) \tag{5.2}$$

$$a_k^p = \sigma(e_k) = \frac{1}{1 + \exp(-e_k)} \tag{5.3}$$

$$Agg\_iECG^p = \sum_{k=1}^{K} a_k^p M_k \tag{5.4}$$

Here, $iECG^p \in \mathbb{R}^{d_M}$ represents the identity ECG vector of person p obtained using the pretrained biometric identification system, and $W, U, V \in \mathbb{R}^{d_M \times d_M}$ denote the learnable parameters of the auxiliary attention network. The attention value e_k^p gauges the similarity between $iECG^p$ and M_k. The attention weights a_k^p are computed by normalizing the attention values e_k^p through a sigmoid activation function. Finally, the Agg_iECG^p vector corresponding to person p is derived using eq 5.4.

$$L_a = \sum_{p=1}^{P} \frac{Agg\_iECG^p \cdot iECG^p}{\|Agg\_iECG^p\| \cdot \|iECG^p\|} \tag{5.5}$$

To ensure the auxiliary attention network effectively learns to derive the Agg_iECG^p vector representing person-specific information, an auxiliary loss (L_a) is introduced. This loss is computed during training based on the cosine similarity between Agg_iECG^p and $iECG^p$, as expressed in eq 5.5. Here, P stands for the total number of subjects in the training dataset. In multi-lead CVD diagnosis scenario, the cumulative auxiliary loss across each lead is computed.

5.1.3 Conditioning Network

The conditioning network estimates the vector parameters, α_n and β_n for the affine transformation of diagnostic features. We propose the utilization of an LSTM network to dynamically compute $\alpha_n(t)$ and $\beta_n(t)$ at each time stamp. This methodology, termed as temporal normalization, facilitates adaptive modulation of diagnostic features, aligning them with the morphological waveform of the ECG signal. The LSTM network's initial hidden state, h_0, is initialized with Agg_iECG^p obtained from the memory module, which ensures the conditioning of the main network's intermediate features is based on person-specific information. At each time stamp t_n, the input to the LSTM network is $Y_n(t_n)$, the output of layer n in the main diagnostic network. The hidden output, denoted as $h_n(t)$, from the LSTM network is then fed into FC layers, producing a scalar value as output. Finally, the output of the FC

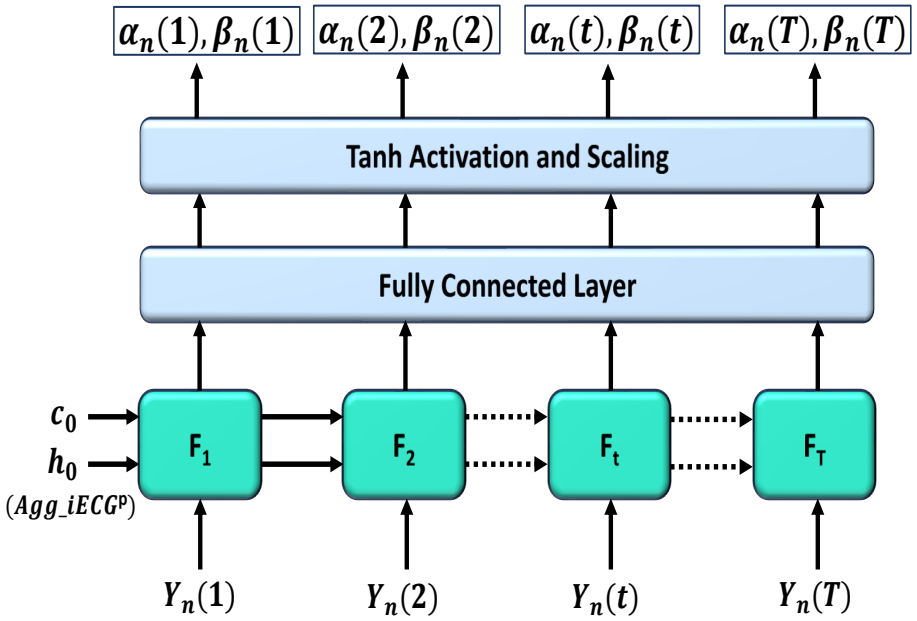

Figure 5.5: Architecture of the Conditioning Network

layer undergoes normalization using the $tanh$ activation function and further scaling. This process is mathematically expressed in equations 5.6 and 5.7.

$$\alpha_n(t) = A \times tanh(W_{c1} h_n(t) + b_{c1}) \tag{5.6}$$

$$\beta_n(t) = B \times tanh(W_{c2} h_n(t) + b_{c2}) \tag{5.7}$$

Here, $W_{c1} \in \mathbb{R}^{d_M \times d_n}$, $b_{c1} \in \mathbb{R}^{d_n}$ and $W_{c2} \in \mathbb{R}^{d_M \times d_n}$, $b_{c2} \in \mathbb{R}^{d_n}$ represent the learnable parameters associated with two fully connected (FC) layers responsible for estimating α_n and β_n. Additionally, A and B denote scaling factors.

5.1.4 Model Optimisation

The proposed person adaptive diagnostic framework is optimised in an end-to-end manner using a joint loss function. The joint loss function is formulated as the summation of the diagnostic loss (L_d) and the auxiliary loss (L_a), as expressed in eq 5.8. We have used a binary cross-entropy loss function for MIT-BIH arrhythmia dataset, since it is formulated as multi-label binary classification scenario. Conversely, for the STAFF-III dataset, L_d is cross-entropy loss function, where the task

involves a multi-class classification problem.

$$L = L_d + L_a \tag{5.8}$$

5.2 Experiment

5.2.1 Database Description

The proposed person-adaptive CVD diagnostic framework is assessed using two publicly available datasets: the MIT-BIH Arrhythmia dataset [193] and the STAFF-III dataset [182]. Detailed descriptions of these datasets are presented in the subsequent subsections.

5.2.1.1 MIT-BIH Arrhythmia Database

The MIT-BIH Arrhythmia database comprises 48 ambulatory ECG recordings, each with two channels and a duration of 30 minutes, extracted from a total of 47 subjects. The ECG signal are sampled at 360 Hz. In this study, we classified 15 distinct types of heartbeats into five categories, adhering to the standards outlined by the Association for the Advancement of Medical Instrumentation (ANSI/AAMI EC57:1998) [196]. The five identified categories encompass normal sinus beats (Normal), supraventricular ectopic beats (SVEB), ventricular ectopic beats (VEB), fusion of a normal and a ventricular ectopic beat (F), and unknown beat type (Q) [137, 196]. For the training phase, we employed 20 recordings (Record IDs starting with Digit 1), and for testing purposes, we allocated 24 recordings(Record IDs starting with Digit 2) [137]. Notably, records with IDs 102, 104, 107, and 217 were excluded due to the presence of paced beats.

5.2.1.2 STAFF-III Database

The STAFF-III dataset contains multi-lead ECG recordings (V1 to V6, Lead I, II, III) from 104 subjects undergoing elective prolonged percutaneous transluminal coronary angiography (PTCA). The dataset comprises both pre-inflation ECG recordings lasting 5 minutes and inflation ECGs with a mean inflation duration of 4 minutes and 23 seconds. In this work, we have excluded patient IDs 1, 19, 48, 89, and 106 due to due to significant corruption caused by noise. The curated dataset comprises 33 subjects with occlusions in the left anterior descending (LAD) artery, 40 with occlusions in the right coronary artery (RCA), and 20 with occlusions in the circumflex artery (CIRC). The training

dataset consists of randomly selected 16, 22, and 10 subjects with blockages in the LAD, RCA, and CIRC artery, respectively. The pre-inflation ECG signal serve as the healthy ECG signals, while the inflation ECG signals are used for respective disease categories.

5.2.2 Evaluation Method

In this work, we have used accuracy (Acc), F1-score (F1), precision (Pre), sensitivity (Sen), specificity (Spe), and threshold free area under the curve (AUC) as the performance metrics. In the case of the multi-label multi-class classification scenario (i.e. MIT-BIH Arrhythmia dataset), we obtained macro-averaged score for better evaluation of model performance.

5.2.3 Implementation Details

5.2.3.1 Network Parameters

The network receives either a single-lead or multi-lead ECG signal, along with an associated $iECG$ vector for each input ECG lead. The dimension of memory vectors (M_k) and $iECG$ vectors, d_M, are configured to 150. The scaling factors A and B are set at 16. This choice enables the model to effectively scale features by a considerable magnitude, fostering improved adaptation and preventing saturation, as emphasized in [190]. The number of clusters, K, in the memory bank is set at 32. The network parameters of the main diagnostic network is as per the parameters proposed in the respective work [114, 174, 184, 189].

5.2.3.2 Training Setting

All deep learning models in this study are implemented using the PyTorch framework. We utilized the Adam optimizer for optimizing the model parameters. The HLSTM-based biometric system is initially trained with a batch size of 100, a learning rate (lr) of 0.01, and 1000 epochs. We selected the pre-trained model with the best test accuracy for subsequent analysis. In the case of training person-adaptive diagnostic models, we meticulously tuned hyperparameters tailored to each diagnostic model. Specifically, we used an initial lr of 0.0001, a batch size of 16, and 150 epochs for ResNet-18 and ResNeXt-50. For HLSTM, Vanilla LSTM, and ASTLNet models, we adopted an initial lr of 0.001, a batch size of 100, and 200 epochs. In all cases, the learning rate is adjusted to (1/5)th of the initial lr using a cosine learning rate scheduler. The experiments were conducted on an NVIDIA P100 GPU facility.

5.2.4 Experiment on MIT-BIH Arrhythmia Datbase

Table 5.1: Performance Comparison of global CVD and person-adaptive CVD diagnosis on the MIT-BIH Arrhythmia dataset

Diagnostic Model	Type	Acc	F1	Pre	Sen	Spe	AUC
Vanilla LSTM [149]	Global	81.94	**38.48**	**47.19**	39.76	87.89	80.84
	Person-Adaptive	**81.95**	38.08	45.46	**39.86**	**88.62**	**81.1**
HLSTM [184]	Global	80.50	36.88	49.41	40.69	88.64	81.53
	Person-Adaptive	**82.41**	**42.06**	**50.04**	**51.42**	**92.16**	**83.84**
ResNet-18 [114]	Global	82.26	40.19	42.53	**47.93**	89.52	81.63
	Person-Adaptive	**83.96**	**40.70**	**48.15**	44.95	**89.81**	**83.37**
ResNeXt-50 [174]	Global	81.38	39.37	45.00	**46.30**	88.61	81.91
	Person-Adaptive	**83.30**	**39.78**	**57.12**	45.98	**89.13**	**82.33**

In this study, we employed channel-1 (Lead-II) of the MIT-BIH Arrhythmia dataset. The iECG vectors associated with each subject are extracted for use as input during both the training and testing phases. For training and testing of person-adaptive diagnostic model, we extracted ECG templates of duration 2 seconds. The training templates are extracted using a rectangular window with an overlap of 75%, while testing templates are extracted without any overlap. The true diagnostic labels for these templates are assigned based on annotations, specifically set 250 ms after the start and 250 ms before the end of each template. During the multi-label diagnostic prediction, a uniform threshold of 0.5 is set for all diagnostic categories. The performance results for both global CVD diagnostic system and person-adaptive diagnostic system are tabulated in Table 5.1. Analysis of Table 5.1 reveals the notable superiority of the person-adaptive diagnostic models across the majority of performance metrics for all the models. Particularly, the person-adaptive models exhibit significant improvements in AUC score across various networks. The person-adaptive HLSTM and ResNet-18 models exhibit an improvement of approximately 2% compared to the global diagnostic model. Similar trends are evident in the accuracy scores for HLSTM, ResNet-18, and ResNeXt-50 models. This shows that the person-adaptive diagnostic model learn to adapt to person-specific information, thereby improving diagnostic performance. The marginal performance enhancement in the case of the Vanilla LSTM model can be attributed to the fact that only the final hidden vector of a single layer undergoes conditioning. In contrast, other models feature multiple layers of conditioning at all timestamps, contributing to their more pronounced improvements. The disease-wise diagnostic performance for three major

categories–Normal, SVEB, and VEB–is comprehensively tabulated in Table 5.3. Table 5.3 reveals better performance in all diagnostic categories for both the HLSTM and ResNet-18 models, showcasing their effectiveness. Additionally, all models demonstrate proficiency in detecting healthy ECG signals.

5.2.5 Experiment on STAFF-III Datbase

Table 5.2: Performance Comparison of global CVD and person-adaptive CVD diagnosis on the STAFF-III dataset

Diagnostic Model	Type	Acc	F1	Pre	Sen	Spe	AUC
Vanilla LSTM [149]	Global	64.27	62.88	67.37	64.34	88.07	85.64
	Person-Adaptive	**66.18**	**64.59**	**68.56**	**64.60**	**88.50**	**85.86**
HLSTM [184]	Global	64.71	**64.33**	66.77	65	88.22	86.41
	Person-Adaptive	**67.91**	64.16	**70.08**	**65**	**89.06**	**87.50**
ResNet-18 [114, 172]	Global	64.51	61.26	64.76	61.67	87.81	84.04
	Person-Adaptive	**67.91**	**64.67**	**66.64**	**65.21**	**89.28**	**85.23**
ResNeXt-50 [174]	Global	64.97	63.10	70.01	63.06	87.68	85.17
	Person-Adaptive	**65.21**	**63.84**	**71.68**	**64.49**	**88.10**	**87.08**
ASTLNet [189]	Global	67.45	64.30	66.97	65.26	89.03	86.87
	Person-Adaptive	**71.61**	**67.16**	**71.1**	**67.77**	**90.14**	**89.39**

The diagnostic performance of both the multi-lead global diagnostic model and the person-adaptive diagnostic model has been assessed and is presented in Table 5.2.The training and testing ECG templates of duration 2s are extracted from each ECG leads similar to the experimental setup of MIT-BIH Arrhythmia dataset. The iECG vectors for all subjects are obtained from each ECG lead using the pre-inflation ECG signal. Table 5.2 demonstrates a substantial enhancement in diagnostic performance with the implementation of person-adaptive diagnostic models compared to their respective global diagnostic counterparts. Specifically, the person-adaptive ASTLNet model demonstrates notable improvements, achieving a diagnostic accuracy increase of over 4% and an AUC score improvement exceeding 2.5%. Likewise, the person-adaptive ResNet-18 model exhibits advancements in accuracy, F1-score, and AUC score by more than 3%, 3%, and 1%, respectively. This validates the effectiveness of the proposed person-adaptive diagnostic framework in learning personalized diagnostic features, leading to significant improvement in diagnostic performance. The confusion matrix associated with each person-adaptive diagnostic model is illustrated in Figure 5.6.

Table 5.3: Performance of the person-adaptive diagnostic models for different disease categories evaluated on the MIT-BIH dataset

Person-Adaptive	Normal						SVEB						VEB						Overall					
Diagnostic Model	Acc	F1	Pre	Sen	Spe	AUC	Acc	F1	Pre	Sen	Spe	AUC	Acc	F1	Pre	Sen	Spe	AUC	Acc	F1	Pre	Sen	Spe	AUC
Vanilla LSTM	93.26	95.91	93.47	92.96	58.75	83.83	94.43	21.10	44.19	15.17	94.36	68.20	92.11	73.38	89.62	71.01	90.32	91.20	81.95	38.08	45.46	39.86	88.62	81.09
HLSTM	93.33	92.34	93.71	87.29	74.15	85.21	93.69	30.35	73.08	66.50	94.33	75.96	91.73	75.38	83.40	82.71	92.75	94.59	82.41	42.06	50.04	51.42	92.16	83.84
Resnet-18	93.09	95.18	94.72	81.71	69.13	82.72	93.87	27.69	71.32	55.08	87.94	74.08	93.11	80.65	74.71	87.95	92.01	97.08	83.96	40.70	48.15	44.95	89.81	83.37
Resnext-50	91.26	94.59	93.49	82.65	66.38	81.88	93.84	32.36	100.00	54.08	90.45	80.95	92.97	71.97	92.10	93.16	88.81	93.77	83.30	39.78	57.12	45.98	89.13	82.33

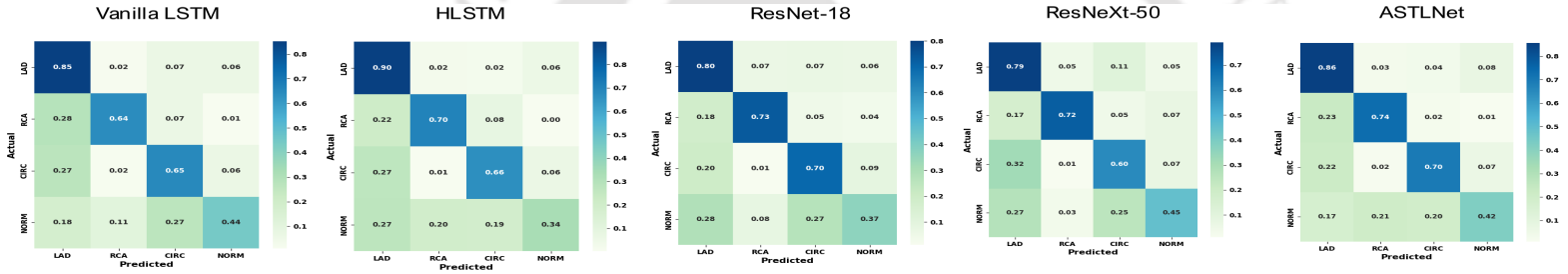

Figure 5.6: Confusion matrix for the person-adaptive diagnostic models evaluated on the STAFF-III dataset

5.2.6 Effect of Temporal Normalization and Multi-layer Normalization

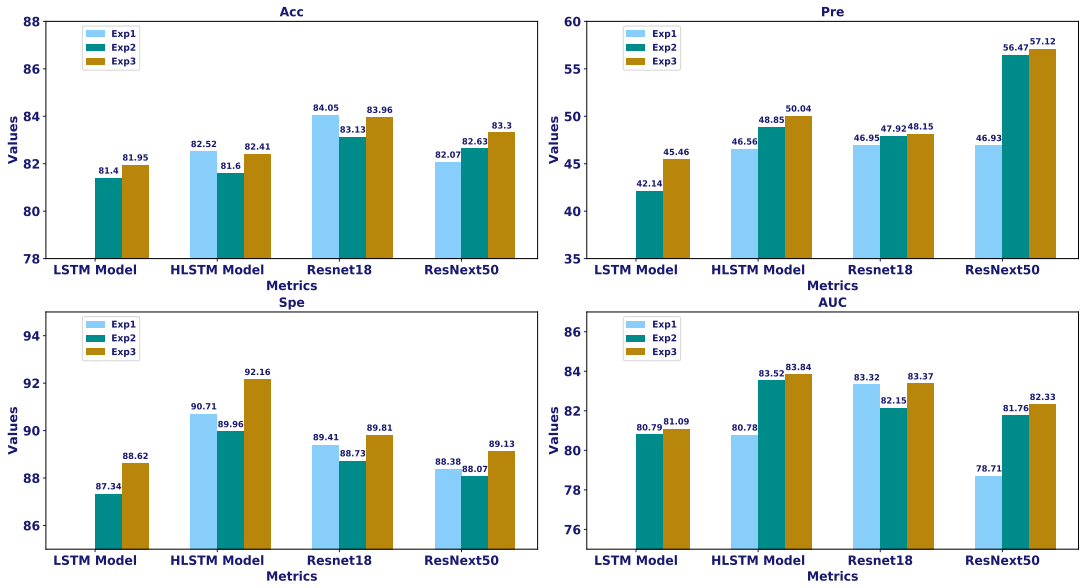

Figure 5.7: Comparison of temporal normalization, all layer normalization and final layer normalization

In this study, we incorporated temporal normalization, enabling the model to dynamically modulate diagnostic features across all timestamps within a layer. Another key aspect of our proposed person-adaptive diagnostic framework involves conditional normalization applied to diagnostic features across all layers. To assess the impact of temporal normalization and multi-layer normalization, we conducted experiments involving conditional normalization without temporal normalization and final layer normalization. The removal of temporal normalization was achieved by introducing an adaptive average pooling layer at the output of the LSTM network in the conditioning network, preceding its input into the FC layer. Figure 5.7 presents a bar plot comparison of the results obtained from three experiments: Exp1, representing final layer normalization; Exp2, representing all-layer normalization without temporal normalization; and Exp3, representing our proposed framework. It's important to note that the Vanilla LSTM network, having only one layer, does not include results for the final layer. The results show Exp3 consistently outperforms Exp1 and Exp2 across most performance metrics. This suggests that temporal normalization, which adaptively incorporates person-specific information into respective morphological waveforms, plays a crucial role. Similarly, infusing person-specific information across all layers can restore person-adaptive diagnostic information across the various layers of the network.

5.2.7 Effect of Number of Clusters in Memory Bank

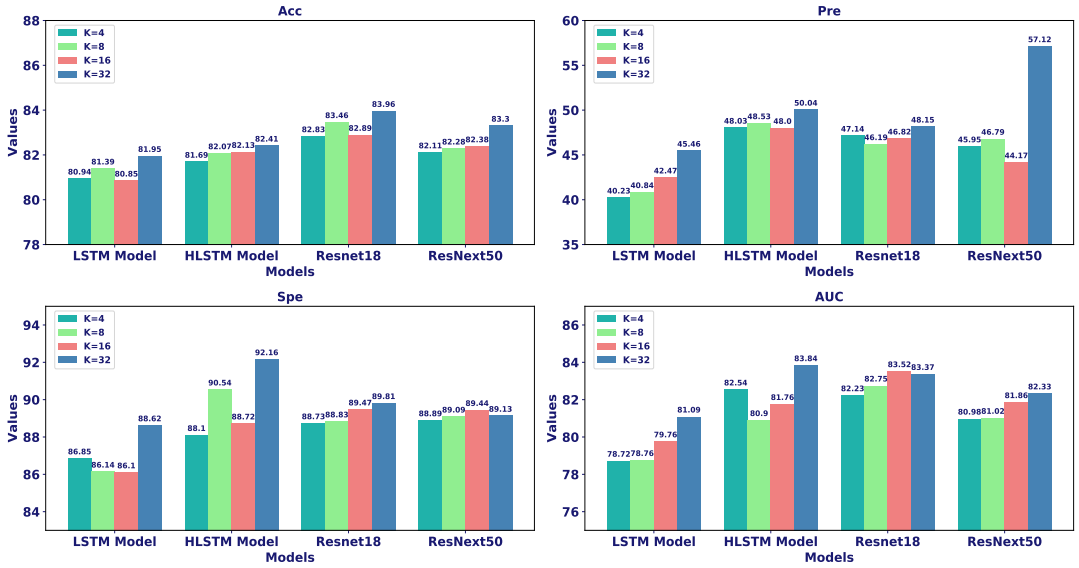

Figure 5.8: Analysis of the effect of number of clusters in memory bank

To assess the influence of the number of clusters in the memory bank on person-adaptive diagnostic feature learning, we conducted experiments with varying cluster counts. The results are illustrated in the bar plot presented in Figure 5.8. The results indicate that the person-adaptive diagnostic model exhibits improved performance with a higher number of cluster centers. This improvement can be attributed to the enhanced representation of person-specific information achieved with an increased number of cluster centers in the memory bank.

5.2.8 Analysis of The Memory Module

The memory module is designed to learn the person-specific information (Agg_iECG) in an unsupervised manner using the memory bank. This allows the model to adapt to a new subject without retraining. To evaluate the effectiveness of the memory module in approximating the person-specific information of a new subject using the memory bank, we computed the cosine similarity score between Agg_iECG^p and $iECG^p$ for a subject (intra-subject), denoted as person p. For comparative analysis, we also computed the cosine similarity of Agg_iECG^p with $iECG^k$, where $k \in 1...P$ and $k \neq p$ (inter-subject). Here, P represents the total number of subjects in the test dataset. The cosine similarity scores were obtained for all persons $p \in 1..P$ and are presented in a box plot in Figure 5.9. The box plot illustrates that the intra-subject score exhibits high values, with the median skewed

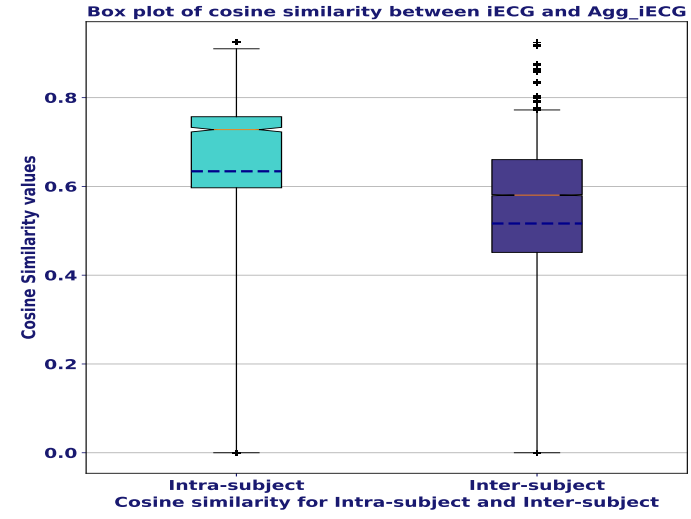

Figure 5.9: Comparison of intra-subject and inter-subject cosine similarity score within iECG and Agg_iECG

towards 1, while the inter-subject score is comparatively lower. This signifies that the memory module effectively learns the person-specific information of a new subject by leveraging the information stored in the memory bank.

5.3 Summary

This chapter presets a novel person-adaptive framework for CVD diagnosis. Specifically, we put forth a novel approach involving the conditional normalization of diagnostic features within a network, based on person-specific information. To capture person-specific information, we incorporated a memory module based on the ECG biometric system. The primary advantage of our proposed framework lies in its modular design and the efficiency of the unsupervised adaptation process. Experimental results affirm the effectiveness of our framework, demonstrating a significant enhancement in diagnostic performance. The proposed framework effectively incorporates person-specific information and adapts diagnostic features in both single and multi-lead ECG recording scenarios.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

6

Conclusions

Contents

6.1 Summary of The Work

The imperative need for application of automated Cardiovascular Disease (CVD) diagnostic system in remote and artificial intelligence based healthcare system has become crucial due to the increasing prevalence and impact of cardiac diseases on public health. However, this progress is accompanied by growing concerns about the security and privacy of sensitive medical data collected during the diagnostic process. The development of a robust automated CVD diagnosis system is particularly challenging due to the tenuous variation in the morphological characteristics of the ECG signal across different diseases. Moreover, there is considerable morphological variability among different categories of subjects, often termed inter-individual variability [3], which leads to performance degradation in the automated diagnostic systems. To tackle these challenges, this thesis work endeavors to develop a secure, automated, person-adaptive CVD diagnosis system. Towards this end, we have proposed two robust ECG based biometric systems that leverage the temporal variation and local morphological shape of the ECG signal. Our investigation revealed that these variations preserve valuable person-specific information. Subsequently, we designed a deep learning model for extracting spatio-tempral variation in multi-lead ECG signals for effective multi-label CVD diagnosis. The person-specific information acquired from the biometric system is then employed in an unsupervised manner to develop a person-adaptive disease diagnosis system.

Chapter 1 presents an insight into the electro-physiological activity of the heart. It offers a concise overview of different ECG acquisition setups, encompassing the gold standard 12-lead ECG recording system, holter recording, and single-lead ECG recording systems. The chapter delves into the morphological characteristics of both healthy ECG signals and those associated with various cardiac abnormalities. Additionally, a comprehensive review of existing ECG-based biometric systems and CVD diagnosis systems is presented in this chapter. The literature review exposes research gaps in the existing works on biometric system, as well as in generalized and person-adaptive diagnostic systems. In the concluding section of this chapter, the research gaps are systematically discussed, thereby establishing a clear motivation for the thesis work.

Chapter 2 presents a novel ECG based biometric system leveraging the temporal variation of the ECG signal. An LSTM based biometric system is proposed to leverage the temporal variation of the

ECG signal for person identification and verification. In this chapter, we advocate the use of ECG segments extracted with a rectangular window of duration $100ms$ and 75% overlap as inputs to the LSTM network. This approach facilitates effective learning of the temporal variation in the ECG signal. A series of experiments are conducted to explore the impact of varying segment lengths on the model's ability to capture the temporal representations. Our findings indicate that smaller ECG segments are particularly beneficial for learning the temporal variation of the ECG signal. However, the basic LSTM network lack in exploiting the multi-scale temporal variation present in the ECG signal. The, multi-scale temporal variation, i.e., intra-beat variation and inter-beat variation of the ECG signal presents significant biometric information. Therefore, an attention-based HLSTM model is designed to capture the multi-scale temporal variation of the ECG signal. An attention mechanism is embedded to the HLSTM model to explicitly focus on specific ECG waveforms that preserve substantial biometric information corresponding to a subject. The HLSTM model is composed of two layers of stacked LSTM network with different update intervals. This hierarchical architecture allows the model to learn the temporal variation of the ECG signal in different abstractions. The major advantage of the proposed method is that it doesn't require the detection of any fiducial points. The proposed framework demonstrated superior performance, utilizing minimal enrollment data–merely 12.5 seconds for each subject and a testing data duration of 2 seconds. The HLSTM-based biometric system proposed in this study exhibited an identification accuracy of 97.4% and 92.1% for the ECG-ID database and CYBHi database, respectively. Notably, the performance of our proposed framework surpasses that of state-of-the-art biometric models.

Permanence of biometric representation extracted from a subject is essential for a robust biometric system. In Chapter 2, we observed that learning temporal representation effectively generates better biometric representation. However, the simultaneous consideration of the multi-scale nature of local morphological waveforms and temporal dynamics is lacking in the current approach to biometric representation learning. To address this gap, we introduce a Multi-Scale Temporal Dynamics Learning Network (MSTDLNet) designed to capture both the local morphological representation and multi-scale temporal dynamics of ECG signals for biometric applications. MSTDLNet exploits the fine-to-coarse flow of information within a stacked convolutional network to learn multi-scale temporal representation. Specifically, we incorporate a convolutional kernel-based Multi-Scale Enhanced Morphological Representation Learning (MSE-MRL) module to enhance the learning of the ECG

waveform's morphological representation. Additionally, we integrate two layers of LSTM networks at different hierarchical levels to innovate and capture the multi-scale temporal dynamics of the ECG signal. A series of experiments are conducted to validate the model performance and analyze the impact of learning multi-scale local morphological shape and temporal variation. Results demonstrate that the biometric representation learned by the model exhibits better permanence property, leading to a significant enhancement in performance during multi-session analysis. The MSTDLNet model achieves an identification accuracy of 98.70% for intra-session analysis and 96.44% for inter-session analysis, as evaluated on the ECG-ID database. Similarly, an identification accuracy of 96.03% for intra-session analysis and 69.84% for inter-session analysis is achieved using the CYBHi database. The incorporation of various attention modules, as proposed in this work, markedly enhances the biometric representation learned using blindly segmented ECG signals. This framework thoroughly explores the multi-scale temporal representation learning method, significantly improving the state-of-the-art in ECG-based biometric systems.

In Chapter 4, we proposed a novel CVD diagnosis system designed to leverage the spatio-temporal variations of the multi-lead ECG signal. The variation of the clinical components of the ECG signal across different leads as well as along the temporal scale constitute the major cue for diagnostic decision making. Therefore, we designed an attentive spatio-temporal learning based neural network (ASTLNet) to effectively capture the concurrent multi-scale spatio-temporal representation from multi-lead ECG signal. This network employs a clustered multi-head criss-cross attention (MHCCA) embedded within a hierarchical LSTM (HLSTM) network to capture the concurrent multi-scale spatio-temporal representation. To effectively aggregate this learned representation, we introduced an attentive spatio-temporal aggregation (ASTA) module. Finally, a multi-aligned attention (MAA) layer was designed and incorporated to derive multiple context vectors, emphasizing multiple diagnostically significant regions in the ECG signal. The proposed model is specifically tailored for multilabel CVD diagnosis applications, enabling the diagnosis of multiple cardiovascular diseases within a single subject. To validate the effectiveness of learning spatio-temporal representation from the ECG signal for CVD diagnosis, we evaluated the model's performance on three publicly available dataset. The results demonstrate that the proposed model performs better with fewer parameters showing the effectiveness of learning spatio-temporal representation for CVD diagnosis.

Finally, an unsupervised person-adaptive CVD diagnosis system using person-specific informa-

tion is presented in Chapter 5. The person-adaptive diagnostic framework incorporates a conditioning network for affine transformation of diagnostic model features based on patient-specific information. Patient-specific details are captured in a knowledge space through a memory module using ECG-based biometric representation (iECG vector). For each new subject, a memory vector is generated by an auxiliary attention network that leverages person-specific information from the knowledge space. An auxiliary loss function is introduced to enhance the learning of a more effective person-specific representation by the memory module. Evaluating the performance of our person-adaptive diagnostic framework, we conducted rigorous assessments on two publicly available ECG databases. The experimental results conclusively affirm the efficacy of our framework, showcasing a noteworthy improvement in diagnostic performance. This improvement in person-adaptive diagnosis is observed in both single and multi-lead ECG recording scenarios, highlighting the versatility and effectiveness of our proposed framework in seamlessly incorporating person-specific information and adapting diagnostic features.

6.2 Scope of The Future Work

In this thesis, we investigated the learning of temporal representation of the ECG signal for biometric application as well as generalized CVD diagnosis. Subsequently, the acquired biometric representations were efficiently employed for person-adaptive CVD diagnosis. Future endeavors in the direction of secure person-adaptive automated CVD diagnosis systems could explore the following directions:

- The ECG signal exhibits variations across different body postures. In Chapter 3, it was observed that the performance of a biometric system tends to decrease when the testing data involves different body postures than the enrollment data. Addressing this challenge is crucial for developing a robust, general-purpose ECG-based biometric system. In this regard, exploration of an unsupervised domain adaptive ECG-based biometric framework is warranted.

- Recording ECG signal in *off-the-person* setup is the most convenient for an ECG-based biometric system. However, the performance tends to decrease when using ECG signals recorded in an off-the-person setup due to low SNR, especially due to those exhibiting high baseline drift and abrupt changes. Biometric identification and verification performance could be enhanced by enrolling and testing subjects with high SNR ECG signals. Therefore, an exploration of a

quality assessment algorithm by estimating the periodicity of the ECG signal can be considered for ECG-based biometric systems.

- We have demonstrated that learning spatio-temporal variation from multi-lead ECG signals enhances diagnostic performance. In the future, researchers could explore various graph neural network architectures explicitly designed to model concurrent spatio-temporal variation. Different attention mechanisms could then be investigated to emphasize diagnostic cues in specific leads.

- The CVD datasets exhibit severe imbalances, with conditions like posterior myocardial infarction (MI), third-degree atrioventricular (AV) block, and atrial flutter being scarce compared to other diseases and healthy signals. To address this issue, various machine learning methods can be investigated with the aim of enhancing sensitivity and precision in detecting low-frequency cardiac diseases while simultaneously improving or maintaining the overall diagnostic performance of the system.

- In this thesis, we investigated the utilization of biometric information to adapt an automated diagnostic system. Future research avenues could involve the exploration of diverse deep generative learning methods to model the various generative factors influencing the variation in the ECG signal. Subsequently, methods aimed at mitigating the impact of these generative factors during diagnosis could be explored, potentially enhancing diagnostic performance.

- Remote monitoring of cardiac health necessitates deploying deep learning-based models on edge devices, which have limited computational power and storage space. However, the proposed frameworks currently require high computational resources. Therefore, future work should explore various quantization and pruning techniques to reduce model complexity, facilitating the deployment of CVD diagnostic models on wearable edge devices. Additionally, integrating various multi-channel signal processing techniques into deep learning-based models can be explored to achieve better performance with lower computational complexity. This integration will also enhance the interpretability of the models, a crucial aspect in healthcare applications.

References

[1] S. S. Virani, A. Alonso, H. J. Aparicio, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, S. Cheng, F. N. Delling *et al.*, "Heart disease and stroke statistics-2021 update: a report from the american heart association." *Circulation*, vol. 143, no. 8, p. CIR0000000000000950, 2021.

[2] Cardiovascular diseases (cvds). Accessed: July 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[3] R. Hoekema, G. J. Uijen, and A. Van Oosterom, "Geometrical aspects of the interindividual variability of multilead ecg recordings," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5, pp. 551–559, 2001.

[4] B. J. Schijvenaars, G. van Herpen, and J. A. Kors, "Intraindividual variability in electrocardiograms," *Journal of Electrocardiology*, vol. 41, no. 3, pp. 190–196, 2008.

[5] S. Geurts, M. J. Tilly, B. Arshi, B. H. Stricker, J. A. Kors, J. W. Deckers, N. M. de Groot, M. A. Ikram, and M. Kavousi, "Heart rate variability and atrial fibrillation in the general population: a longitudinal and mendelian randomization study," *Clinical Research in Cardiology*, vol. 112, no. 6, pp. 747–758, 2023.

[6] G. Wang, M. Chen, Z. Ding, J. Li, H. Yang, and P. Zhang, "Inter-patient ecg arrhythmia heartbeat classification based on unsupervised domain adaptation," *Neurocomputing*, vol. 454, pp. 339–349, 2021.

[7] P. K. Gyawali, J. V. Murkute, M. Toloubidokhti, X. Jiang, B. M. Horacek, J. L. Sapp, and L. Wang, "Learning to disentangle inter-subject anatomical variations in electrocardiographic data," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 2, pp. 860–870, 2021.

[8] M. Porumb, S. Stranges, A. Pescapè, and L. Pecchia, "Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ecg," *Scientific reports*, vol. 10, no. 1, p. 170, 2020.

[9] E. M. Antman and J. Loscalzo, "Precision medicine in cardiology," *Nature Reviews Cardiology*, vol. 13, no. 10, p. 591, 2016.

[10] H. P. Da Silva, A. Lourenço, A. Fred, N. Raposo, and M. Aires-de Sousa, "Check your biosignals here: A new dataset for off-the-person ecg biometrics," *Computer methods and programs in biomedicine*, vol. 113, no. 2, pp. 503–514, 2014.

[11] W. B. Fye, "A history of the origin, evolution, and impact of electrocardiography," *The American journal of cardiology*, vol. 73, no. 13, pp. 937–949, 1994.

[12] E. Frank, "An accurate, clinically practical system for spatial vectorcardiography," *circulation*, vol. 13, no. 5, pp. 737–749, 1956.

[13] A. H. Association *et al.*, *Standardization of Precordial Leads*. The Association., 1938.

[14] E. Goldberger, "A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar, extremity leads," *American Heart Journal*, vol. 23, no. 4, pp. 483–492, 1942.

[15] R. E. Mason and I. Likar, "A new system of multiple-lead exercise electrocardiography," *American heart journal*, vol. 71, no. 2, pp. 196–205, 1966.

[16] A. D. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *IEEE transactions on instrumentation and measurement*, vol. 57, no. 2, pp. 248–253, 2008.

[17] G. G. Molina, F. Bruekers, C. Presura, M. Damstra, and M. van der Veen, "Morphological synthesis of ecg signals for person authentication," in *2007 15th European Signal Processing Conference*. IEEE, 2007, pp. 738–742.

[18] Q. Zhang, D. Zhou, and X. Zeng, "Pulseprint: Single-arm-ecg biometric human identification using deep learning," in *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. IEEE, 2017, pp. 452–456.

[19] D. P. Coutinho, A. L. Fred, and M. A. Figueiredo, "One-lead ecg-based personal identification using ziv-merhav cross parsing," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 3858–3861.

[20] F. N. Wilson, A. G. Macleod, and P. S. Barker, "The potential variations produced by the heart beat at the apices of einthoven's triangle," *American Heart Journal*, vol. 7, no. 2, pp. 207–211, 1931.

[21] A. Goldberger, *Goldberger's Clinical Electrocardiography*. Elsevier, 2018.

[22] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "Ecg analysis: a new approach in human identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 3, pp. 808–812, 2001.

[23] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Evolution, current challenges, and future possibilities in ecg biometrics," *IEEE Access*, vol. 6, pp. 34 746–34 776, 2018.

[24] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh, "Ecg biometric recognition: A comparative analysis," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1812–1824, 2012.

[25] A. N. Uwaechia and D. A. Ramli, "A comprehensive survey on ecg signals as new biometric modality for human authentication: Recent advances and future challenges," *IEEE Access*, vol. 9, pp. 97 760–97 802, 2021.

[26] R. D. Labati, E. Muñoz, V. Piuri, R. Sassi, and F. Scotti, "Deep-ecg: Convolutional neural networks for ecg biometric recognition," *Pattern Recognition Letters*, vol. 126, pp. 78–85, 2019.

[27] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "Human identification from ecg signals via sparse representation of local segments," *IEEE Signal Processing Letters*, vol. 20, no. 10, pp. 937–940, 2013.

[28] S. S. Abdeldayem and T. Bourlai, "A novel approach for ecg-based human identification using spectral correlation and deep learning," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2019.

[29] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, "On evaluating ecg biometric systems: session-dependence and body posture," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 2002–2013, 2014.

[30] W. Louis, M. Komeili, and D. Hatzinakos, "Continuous authentication using one-dimensional multi-resolution local binary patterns (1dmrlbp) in ecg biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2818–2832, 2016.

[31] E. J. da Silva Luz, G. J. Moreira, L. S. Oliveira, W. R. Schwartz, and D. Menotti, "Learning deep off-the-person heart biometrics representations," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1258–1270, 2017.

[32] Y. Huang, G. Yang, K. Wang, H. Liu, and Y. Yin, "Learning joint and specific patterns: A unified sparse representation for off-the-person ecg biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 147–160, 2020.

[33] G. Zhu, M. Ma, Y. Huang, K. Wang, and G. Yang, "Dual-domain low-rank fusion deep metric learning for off-the-person ecg biometrics," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 2914–2918.

[34] R. M. Rangayyan, *Biomedical signal analysis*. John Wiley & Sons, 2015, vol. 33.

[35] F. Agrafioti, F. M. Bui, D. Hatzinakos *et al.*, "Secure telemedicine: Biometrics for remote and continuous patient verification," *Journal of Computer Networks and Communications*, vol. 2012, 2012.

[36] M. Komeili, W. Louis, N. Armanfard, and D. Hatzinakos, "Feature selection for nonstationary data: Application to human recognition using medical biometrics," *IEEE Transactions on Cybernetics*, vol. 48, no. 5, pp. 1446–1459, 2017.

[37] P. De Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O'Malley, "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea," *IEEE transactions on biomedical engineering*, vol. 50, no. 6, pp. 686–696, 2003.

[38] L. Sharma, S. Dandapat, and A. Mahanta, "Ecg signal denoising using higher order statistics in wavelet subbands," *Biomedical Signal Processing and Control*, vol. 5, no. 3, pp. 214–222, 2010.

[39] S. R. Krishnan and C. S. Seelamantula, "On the selection of optimum savitzky-golay filters," *IEEE transactions on signal processing*, vol. 61, no. 2, pp. 380–391, 2012.

[40] R. Salloum and C.-C. J. Kuo, "Ecg-based biometrics using recurrent neural networks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2062–2066.

[41] C. L. P. Lim, W. L. Woo, S. S. Dlay, and B. Gao, "Heartrate-dependent heartwave biometric identification with thresholding-based gmm–hmm methodology," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 45–53, 2018.

[42] S.-C. Wu, P.-T. Chen, A. L. Swindlehurst, and P.-L. Hung, "Cancelable biometric recognition with ecgs: subspace-based approaches," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1323–1336, 2018.

[43] S.-C. Wu, P.-L. Hung, and A. L. Swindlehurst, "Ecg biometric recognition: Unlinkability, irreversibility and security," *IEEE Internet of Things Journal*, 2020.

[44] K. A. Sidek, I. Khalil, and H. F. Jelinek, "Ecg biometric with abnormal cardiac conditions in remote monitoring system," *IEEE Transactions on systems, man, and cybernetics: systems*, vol. 44, no. 11, pp. 1498–1509, 2014.

[45] J. M. Irvine, S. A. Israel, W. T. Scruggs, and W. J. Worek, "eigenpulse: Robust human identification from cardiovascular function," *Pattern Recognition*, vol. 41, no. 11, pp. 3427–3435, 2008.

[46] Q. Zhang, D. Zhou, and X. Zeng, "Heartid: A multiresolution convolutional neural network for ecg-based biometric human identification in smart health applications," *Ieee Access*, vol. 5, pp. 11 805–11 816, 2017.

[47] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "Ecg to identify individuals," *Pattern recognition*, vol. 38, no. 1, pp. 133–142, 2005.

[48] J. S. Arteaga-Falconi, H. Al Osman, and A. El Saddik, "Ecg authentication for mobile devices," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 591–600, 2015.

[49] P. Huang, L. Guo, M. Li, and Y. Fang, "Practical privacy-preserving ecg-based authentication for iot-based healthcare," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9200–9210, 2019.

[50] R. Cordeiro, D. Gajaria, A. Limaye, T. Adegbija, N. Karimian, and F. Tehranipoor, "Ecg-based authentication using timing-aware domain-specific architecture," *IEEE transactions on computer-aided design of integrated circuits and systems*, vol. 39, no. 11, pp. 3373–3384, 2020.

[51] K. Wang, G. Yang, Y. Huang, and Y. Yin, "Multi-scale differential feature for ecg biometrics with collective matrix factorization," *Pattern Recognition*, vol. 102, p. 107211, 2020.

[52] M. Li and S. Narayanan, "Robust ecg biometrics by fusing temporal and cepstral information," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 1326–1329.

[53] J. Liu, L. Yin, C. He, B. Wen, X. Hong, and Y. Li, "A multiscale autoregressive model-based electrocardiogram identification method," *IEEE Access*, vol. 6, pp. 18 251–18 263, 2018.

[54] K. N. Plataniotis, D. Hatzinakos, and J. K. Lee, "Ecg biometric recognition without fiducial detection," in *2006 Biometrics symposium: Special session on research at the biometric consortium conference*. IEEE, 2006, pp. 1–6.

[55] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP journal on Advances in Signal Processing*, vol. 2008, no. 1, p. 148658, 2007.

[56] F. Agrafioti and D. Hatzinakos, "Fusion of ecg sources for human identification," in *2008 3rd International Symposium on Communications, Control and Signal Processing*. IEEE, 2008, pp. 1542–1547.

[57] S.-C. Fang and H.-L. Chan, "Human identification by quantifying similarity and dissimilarity in electrocardiogram phase space," *Pattern Recognition*, vol. 42, no. 9, pp. 1824–1831, 2009.

[58] J.-N. Lee and K.-C. Kwak, "Personal identification using a robust eigen ecg network based on time-frequency representations of ecg signals," *IEEE access*, vol. 7, pp. 48 392–48 404, 2019.

[59] A. Goshvarpour and A. Goshvarpour, "Human identification using a new matching pursuit-based feature set of ecg," *Computer methods and programs in biomedicine*, vol. 172, pp. 87–94, 2019.

[60] Z. Zhao and L. Yang, "Ecg identification based on matching pursuit," in *2011 4th International conference on biomedical engineering and informatics (BMEI)*, vol. 2. IEEE, 2011, pp. 721–724.

[61] R. Li, G. Yang, K. Wang, Y. Huang, F. Yuan, and Y. Yin, "Robust ecg biometrics using gnmf and sparse representation," *Pattern Recognition Letters*, vol. 129, pp. 70–76, 2020.

[62] C. Tan, L. Zhang, T. Qian, S. Brás, and A. J. Pinho, "Statistical n-best afd-based sparse representation for ecg biometric identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[63] K. Wang, G. Yang, Y. Huang, L. Yang, and Y. Yin, "Online ecg biometrics via hadamard code," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 2924–2928.

[64] Y. Zhang, Z. Xiao, Z. Guo, and Z. Wang, "Ecg-based personal recognition using a convolutional neural network," *Pattern Recognition Letters*, vol. 125, pp. 668–676, 2019.

[65] Y. Li, Y. Pang, K. Wang, and X. Li, "Toward improving ecg biometric identification using cascaded convolutional neural networks," *Neurocomputing*, vol. 391, pp. 83–95, 2020.

[66] R. Srivastva, A. Singh, and Y. N. Singh, "Plexnet: A fast and robust ecg biometric system for human recognition," *Information Sciences*, vol. 558, pp. 208–228, 2021.

[67] J. R. Pinto, M. V. Correia, and J. S. Cardoso, "Secure triplet loss: Achieving cancelability and non-linkability in end-to-end deep biometrics," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 2, pp. 180–189, 2020.

[68] M. Sepahvand and F. Abdali-Mohammadi, "A novel multi-lead ecg personal recognition based on signals functional and structural dependencies using time-frequency representation and evolutionary morphological cnn," *Biomedical Signal Processing and Control*, vol. 68, p. 102766, 2021.

[69] Z. Zhao, Y. Zhang, Y. Deng, and X. Zhang, "Ecg authentication system design incorporating a convolutional neural network and generalized s-transformation," *Computers in biology and medicine*, vol. 102, pp. 168–179, 2018.

[70] M. Hammad, S. Zhang, and K. Wang, "A novel two-dimensional ecg feature extraction and classification algorithm based on convolution neural network for human authentication," *Future Generation Computer Systems*, vol. 101, pp. 180–196, 2019.

[71] L. Sun, Z. Zhong, Z. Qu, and N. Xiong, "Perae: An effective personalized autoencoder for ecg-based biometric in augmented reality system," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 6, pp. 2435–2446, 2022.

[72] P.-Y. Hsu, P.-H. Hsu, and H.-L. Liu, "Fold electrocardiogram into a fingerprint," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 828–829.

[73] H. M. Lynn, S. B. Pan, and P. Kim, "A deep bidirectional gru network model for biometric electrocardiogram classification based on recurrent neural networks," *IEEE Access*, vol. 7, pp. 145 395–145 405, 2019.

[74] A. J. Prakash, K. K. Patro, M. Hammad, R. Tadeusiewicz, and P. Pławiak, "Baed: A secured biometric authentication system using ecg signal based on deep learning techniques," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 4, pp. 1081–1093, 2022.

[75] S. K. Cherupally, S. Yin, D. Kadetotad, G. Srivastava, C. Bae, S. J. Kim, and J.-s. Seo, "Ecg authentication hardware design with low-power signal processing and neural network optimization with low precision and structured compression," *IEEE transactions on biomedical circuits and systems*, vol. 14, no. 2, pp. 198–208, 2020.

[76] U. Satija, B. Ramkumar, and M. S. Manikandan, "A review of signal processing techniques for electrocardiogram signal quality assessment," *IEEE reviews in biomedical engineering*, vol. 11, pp. 36–52, 2018.

[77] ——, "Automated ecg noise detection and classification system for unsupervised healthcare monitoring," *IEEE Journal of biomedical and health informatics*, vol. 22, no. 3, pp. 722–732, 2017.

[78] A. Rizwan, A. Zoha, I. B. Mabrouk, H. M. Sabbour, A. S. Al-Sumaiti, A. Alomainy, M. A. Imran, and Q. H. Abbasi, "A review on the state of the art in atrial fibrillation detection enabled by machine learning," *IEEE reviews in biomedical engineering*, vol. 14, pp. 219–239, 2020.

[79] 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.

[80] S. Ansari et al., "A review of automated methods for detection of myocardial ischemia and infarction using electrocardiogram and electronic health records," *IEEE Rev. Biomed. Eng.*, vol. 10, pp. 264–298, 2017.

[81] X. Liu, H. Wang, Z. Li, and L. Qin, "Deep learning in ecg diagnosis: A review," *Knowledge-Based Systems*, vol. 227, p. 107187, 2021.

[82] J. Fayn, "A classification tree approach for cardiac ischemia detection using spatiotemporal information from three standard ecg leads," *IEEE. Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 95–102, 2010.

[83] L. Sharma, R. Tripathy, and S. Dandapat, "Multiscale energy and eigenspace approach to detection and localization of myocardial infarction," *IEEE. Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1827–1837, 2015.

[84] L. Sun et al., "Ecg analysis using multiple instance learning for myocardial infarction detection," *IEEE. Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3348–3356, 2012.

[85] F. Alonso-Atienza et al., "Detection of life-threatening arrhythmias using feature selection and support vector machines," *IEEE. Trans. Biomed. Eng.*, vol. 61, no. 3, pp. 832–840, 2013.

[86] M. Arif, I. A. Malagore, and F. A. Afsar, "Detection and localization of myocardial infarction using k-nearest neighbor classifier," *J. Med. Syst.*, vol. 36, no. 1, pp. 279–289, 2012.

[87] R. Andreao, B. Dorizzi, J. Boudy, and J. Mota, "St-segment analysis using hidden markov model beat segmentation: application to ischemia detection," in *Computers in Cardiology, 2004.* IEEE, 2004, pp. 381–384.

[88] B.-H. Kung et al., "An efficient ecg classification system using resource-saving architecture and random forest," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 6, pp. 1904–1914, 2020.

[89] E. Pueyo, L. Sornmo, and P. Laguna, "Qrs slopes for detection and characterization of myocardial ischemia," *IEEE transactions on Biomedical Engineering*, vol. 55, no. 2, pp. 468–477, 2008.

[90] B. Doğan and M. Korürek, "A new ecg beat clustering method based on kernelized fuzzy c-means and hybrid ant colony optimization for continuous domains," *Applied Soft Computing*, vol. 12, no. 11, pp. 3442–3451, 2012.

[91] X. Tang, Z. Ma, Q. Hu, and W. Tang, "A real-time arrhythmia heartbeats classification algorithm using parallel delta modulations and rotated linear-kernel support vector machines," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 4, pp. 978–986, 2019.

[92] C. Ye, B. V. Kumar, and M. T. Coimbra, "Heartbeat classification using morphological and dynamic features of ecg signals," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2930–2941, 2012.

[93] S. Raj and K. C. Ray, "Ecg signal analysis using dct-based dost and pso optimized svm," *IEEE Transactions on instrumentation and measurement*, vol. 66, no. 3, pp. 470–478, 2017.

[94] M. Stridh, L. Sornmo, C. J. Meurling, and S. B. Olsson, "Sequential characterization of atrial tachyarrhythmias based on ecg time-frequency analysis," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 1, pp. 100–114, 2004.

[95] E. Valverde and P. Arini, "Study of t-wave spectral variance during acute myocardial ischemia," in *2012 Computing in Cardiology*. IEEE, 2012, pp. 653–656.

[96] B. Liu, J. Liu, G. Wang, K. Huang, F. Li, Y. Zheng, Y. Luo, and F. Zhou, "A novel electrocardiogram parameterization algorithm and its application in myocardial infarction detection," *Computers in biology and medicine*, vol. 61, pp. 178–184, 2015.

[97] B. Fatimah, P. Singh, A. Singhal, D. Pramanick, S. Pranav, and R. B. Pachori, "Efficient detection of myocardial infarction from single lead ecg signal," *Biomedical Signal Processing and Control*, vol. 68, p. 102678, 2021.

[98] S. I. Khan and R. B. Pachori, "Derived vectorcardiogram based automated detection of posterior myocardial infarction using fbse-ewt technique," *Biomedical Signal Processing and Control*, vol. 70, p. 103051, 2021.

[99] E. Pasolli and F. Melgani, "Active learning methods for electrocardiographic signal classification," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 6, pp. 1405–1416, 2010.

[100] N. Sinha and A. Das, "Identification and localization of myocardial infarction based on analysis of ecg signal in cross spectral domain using boosted svm classifier," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[101] M. Abdelazez, S. Rajan, and A. D. Chan, "Detection of atrial fibrillation in compressively sensed electrocardiogram measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2020.

[102] S. Raj and K. C. Ray, "Sparse representation of ecg signals for automated recognition of cardiac arrhythmias," *Expert systems with applications*, vol. 105, pp. 49–64, 2018.

[103] P.-C. Chang et al., "Myocardial infarction classification with multi-lead ecg using hidden markov models and gaussian mixture models," *Appl. Soft Comput.*, vol. 12, no. 10, pp. 3165–3175, 2012.

[104] D. A. Coast, R. M. Stern, G. G. Cano, and S. A. Briller, "An approach to cardiac arrhythmia analysis using hidden markov models," *IEEE Transactions on biomedical Engineering*, vol. 37, no. 9, pp. 826–836, 1990.

[105] R. V. Andreao, B. Dorizzi, and J. Boudy, "Ecg signal analysis through hidden markov models," *IEEE. Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1541–1549, 2006.

[106] J. Cheng, L. Dong, and M. Lapata, "Long short-term memory-networks for machine reading," *arXiv preprint arXiv:1601.06733*, 2016.

[107] S. Padhy, "Multilead ecg data analysis using svd and higher-order svd," Ph.D. dissertation, Ph. D. thesis, Indian Institute of Technology Guwahati, India, 2017.

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

[109] 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.

[110] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ecg features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 48, no. 12, pp. 2095–2104, 2018.

[111] U. B. Baloglu et al., "Classification of myocardial infarction with multi-lead ecg signals and deep cnn," *Pattern Recognit. Lett.*, vol. 122, pp. 23–30, 2019.

[112] Y. Cao et al., "Ml-net: Multi-channel lightweight network for detecting myocardial infarction," *IEEE J. Biomed. Health Inform.*, 2021.

[113] C. Han and L. Shi, "Ml–resnet: A novel network to detect and locate myocardial infarction using 12 leads ecg," *Comput. Methods. Programs. Biomed.*, vol. 185, p. 105138, 2020.

[114] D. Zhang et al., "Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram," *Iscience*, vol. 24, no. 4, p. 102373, 2021.

[115] 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.

[116] 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.

[117] L. Qin et al., "An end-to-end 12-leading electrocardiogram diagnosis system based on deformable convolutional neural network with good antinoise ability," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[118] P. Ivaturi et al., "A comprehensive explanation framework for biomedical time series classification," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2398–2408, 2021.

[119] Q. Yao et al., "Multi-class arrhythmia detection from 12-lead varied-length ecg using attention-based time-incremental convolutional neural network," *Inf. Fusion*, vol. 53, pp. 174–182, 2020.

[120] W. Liu et al., "Mfb-cbrnn: A hybrid network for mi detection using 12-lead ecgs," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 503–514, 2020.

[121] M. Hammad, A. M. Iliyasu, A. Subasi, E. S. Ho, and A. A. Abd El-Latif, "A multitier deep learning model for arrhythmia detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2020.

[122] S. Saadatnejad, M. Oveisi, and M. Hashemi, "Lstm-based ecg classification for continuous monitoring on personal wearable devices," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 2, pp. 515–523, 2019.

[123] B. Hou et al., "Lstm-based auto-encoder model for ecg arrhythmias classification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1232–1240, 2019.

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

[125] Y. Gao, H. Wang, and Z. Liu, "A novel approach for atrial fibrillation signal identification based on temporal attention mechanism," in *2020 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.* IEEE, 2020, pp. 316–319.

[126] Y. Jin et al., "A novel interpretable method based on dual level attentional deep neural network for actual multi label arrhythmia detection," *IEEE Trans. Instrum. Meas.*, 2021.

[127] J. Yoo, T. J. Jun, and Y.-H. Kim, "xecgnet: Fine-tuning attention map within convolutional neural network to improve detection and explainability of concurrent cardiac arrhythmias," *Comput. Methods. Programs. Biomed.*, vol. 208, p. 106281, 2021.

[128] J. Wang et al., "Automated ecg classification using a non-local convolutional block attention module," *Comput. Methods. Programs. Biomed.*, vol. 203, p. 106006, 2021.

[129] G. W. Colopy, S. J. Roberts, and D. A. Clifton, "Bayesian optimization of personalized models for patient vital-sign monitoring," *IEEE journal of biomedical and health informatics*, vol. 22, no. 2, pp. 301–310, 2017.

[130] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ecg beat classifier using a mixture of experts approach," *IEEE transactions on biomedical engineering*, vol. 44, no. 9, pp. 891–900, 1997.

[131] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ecg signals," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1415–1426, 2009.

[132] M. Llamedo and J. P. Martínez, "An automatic patient-adapted ecg heartbeat classifier allowing expert assistance," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2312–2320, 2012.

[133] C. Ye, B. V. Kumar, and M. T. Coimbra, "An automatic subject-adaptable heartbeat classifier based on multiview learning," *IEEE journal of biomedical and health informatics*, vol. 20, no. 6, pp. 1485–1492, 2015.

[134] Z. Zhou, X. Zhai, and C. Tin, "Fully automatic electrocardiogram classification system based on generative adversarial network with auxiliary classifier," *Expert Systems with Applications*, vol. 174, p. 114809, 2021.

[135] R. Watrous and G. Towell, "A patient-adaptive neural network ecg patient monitoring algorithm," in *Computers in Cardiology 1995.* IEEE, 1995, pp. 229–232.

[136] P. De Chazal and R. B. Reilly, "A patient-adapting heartbeat classifier using ecg morphology and heartbeat interval features," *IEEE transactions on biomedical engineering*, vol. 53, no. 12, pp. 2535–2543, 2006.

[137] W. Jiang and S. G. Kong, "Block-based neural networks for personalized ecg signal classification," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1750–1761, 2007.

[138] 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.

[139] P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T.-s. Zhou, T. Li, and J.-s. Li, "High-performance personalized heartbeat classification model for long-term ecg signal," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 78–86, 2016.

[140] S. S. Xu, M.-W. Mak, and C.-C. Cheung, "I-vector-based patient adaptation of deep neural networks for automatic heartbeat classification," *IEEE journal of biomedical and health informatics*, vol. 24, no. 3, pp. 717–727, 2019.

[141] T. Golany and K. Radinsky, "Pgans: Personalized generative adversarial networks for ecg synthesis to improve patient-specific deep ecg classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 557–564.

[142] X. Zhai, Z. Zhou, and C. Tin, "Semi-supervised learning for ecg classification without patient-specific labeled data," *Expert Systems with Applications*, vol. 158, p. 113411, 2020.

[143] M. Yamaç, M. Duman, İ. Adalıoğlu, S. Kiranyaz, and M. Gabbouj, "A personalized zero-shot ecg arrhythmia monitoring system: From sparse representation based domain adaption to energy efficient abnormal beat detection for practical ecg surveillance," *arXiv preprint arXiv:2207.07089*, 2022.

[144] P. K. Gyawali, B. M. Horacek, J. L. Sapp, and L. Wang, "Sequential factorized autoencoder for localizing the origin of ventricular activation from 12-lead electrocardiograms," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 5, pp. 1505–1516, 2019.

[145] A. K. Jain, A. Ross, S. Prabhakar *et al.*, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, 2004.

[146] S. Pirbhulal, H. Zhang, W. Wu, S. C. Mukhopadhyay, and Y.-T. Zhang, "Heartbeats based biometric random binary sequences generation to secure wireless body sensor networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 12, pp. 2751–2759, 2018.

[147] S. M. Qaisar and A. Subasi, "Cloud-based ecg monitoring using event-driven ecg acquisition and machine learning techniques," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 623–634, 2020.

[148] Y. Sun, K. L. Chan, and S. M. Krishnan, "Ecg signal conditioning by morphological filtering," *Computers in biology and medicine*, vol. 32, no. 6, pp. 465–479, 2002.

[149] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

[150] V. Campos, B. Jou, X. G. i Nieto, J. Torres, and S.-F. Chang, "Skip RNN: Learning to skip state updates in recurrent neural networks," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=HkwVAXyCW

[151] A. G. A. P. Goyal, A. Sordoni, M.-A. Côté, N. R. Ke, and Y. Bengio, "Z-forcing: Training stochastic recurrent networks," in *Advances in neural information processing systems*, 2017, pp. 6713–6723.

[152] G. Melis, T. Kočiský, and P. Blunsom, "Mogrifier lstm," in *International Conference on Learning Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=SJe5P6EYvS

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

[154] T. S. Lugovaya, "Biometric human identification based on ecg," *PhysioNet*, 2005.

[155] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[156] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet," *Biomedizinische Technik/Biomedical Engineering*, vol. 40, no. s1, pp. 317–318, 1995.

[157] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, pp. 8024–8035.

[158] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[159] 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*, 2015, pp. 2048–2057.

[160] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[161] M. N. Dar, M. U. Akram, A. Usman, and S. A. Khan, "Ecg biometric identification for general population using multiresolution analysis of dwt based features," in *2015 Second International Conference on Information Security and Cyber Forensics (InfoSec)*. IEEE, 2015, pp. 5–10.

[162] J. Gao, Z. Shi, G. Wang, J. Li, Y. Yuan, S. Ge, and X. Zhou, "Accurate temporal action proposal generation with relation-aware pyramid network," in *Proc. Innov. Appl. Artif. Intell. Conf.*, vol. 34, no. 07, 2020, pp. 10 810–10 817.

[163] K. Bandara, C. Bergmeir, and H. Hewamalage, "Lstm-msnet: Leveraging forecasts on sets of related time series with multiple seasonal patterns," *IEEE Trans. Neural. Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1586–1599, 2020.

[164] Q. Ma, Z. Lin, E. Chen, and G. Cottrell, "Temporal pyramid recurrent neural network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 5061–5068.

[165] D. Shan, Y. Luo, X. Zhang, and C. Zhang, "Drrnets: Dynamic recurrent routing via low-rank regularization in recurrent neural networks," *IEEE Trans. Neural. Netw. Learn. Syst.*, 2021.

[166] C. Beyan, S. Karumuri, G. Volpe, A. Camurri, and R. Niewiadomski, "Modeling multiple temporal scales of full-body movements for emotion classification," *IEEE Transactions on Affective Computing*, 2021.

[167] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 2, pp. 652–662, 2019.

[168] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," *arXiv preprint arXiv:1611.01144*, 2016.

[169] C. J. Maddison, A. Mnih, and Y. W. Teh, "The concrete distribution: A continuous relaxation of discrete random variables," *arXiv preprint arXiv:1611.00712*, 2016.

[170] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[171] D. Jyotishi and S. Dandapat, "An ecg biometric system using hierarchical lstm with attention mechanism," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 6052–6061, 2021.

[172] 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.

[173] D. Jyotishi and S. Dandapat, "A multi-scale residual neural network for ecg based person identification," in *2022 IEEE 19th India Council Int. Conf. (INDICON)*. IEEE, 2022, pp. 1–6.

[174] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1492–1500.

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

[176] M. Nishiga et al., "Covid-19 and cardiovascular disease: from basic mechanisms to clinical perspectives," *Nat. Rev. Cardiol.*, vol. 17, no. 9, pp. 543–558, 2020.

[177] S. Hong et al., "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review," *Comput. Biol. Med.*, vol. 122, p. 103801, 2020.

[178] N. Strodthoff et al., "Deep learning for ecg analysis: Benchmarks and insights from ptb-xl," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 5, pp. 1519–1528, 2020.

[179] Z. Huang et al., "Ccnet: Criss-cross attention for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 603–612.

[180] D. Jyotishi and S. Dandapat, "An lstm based model for person identification using ecg signal," *IEEE Sensors Letters*, 2020.

[181] P. Wagner et al., "Ptb-xl, a large publicly available electrocardiography dataset," *Sci. Data*, vol. 7, no. 1, pp. 1–15, 2020.

[182] F. Liu 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.

[183] W. Liu et al., "Real-time multilead convolutional neural network for myocardial infarction detection," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1434–1444, 2017.

[184] D. Jyotishi and S. Dandapat, "An attention based hierarchical lstm model for detection of myocardial infarction," in *2020 IEEE 17th India Counc. Int. Conf.* IEEE, 2020, pp. 1–5.

[185] L. D. Sharma and R. K. Sunkaria, "Inferior myocardial infarction detection using stationary wavelet transform and machine learning approach," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 199–206, 2018.

[186] T. Reasat and C. Shahnaz, "Detection of inferior myocardial infarction using shallow convolutional neural networks," in *2017 IEEE R10-HTC*. IEEE, 2017, pp. 718–721.

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

[188] 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.

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

[190] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[191] D. Ha, A. Dai, and Q. Le, "Hypernetworks," 2016.

[192] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," in *International conference on machine learning*. PMLR, 2017, pp. 1243–1252.

[193] 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.

[194] Y. Zhao, C. Ni, C.-C. Leung, S. R. Joty, E. S. Chng, and B. Ma, "Speech transformer with speaker aware persistent memory." in *INTERSPEECH*, 2020, pp. 1261–1265.

[195] J. Pan, G. Wan, J. Du, and Z. Ye, "Online speaker adaptation using memory-aware networks for speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1025–1037, 2020.

[196] E. Ansi-Aami, "Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms assoc. adv," *Med. Instrum., Arlington, VA*, 1998.

List of Publications

Journal Publications

- Published Paper and Accepted Publication:

  1. **D. Jyotishi** and S. Dandapat, “An LSTM-based model for person identification using ECG signal”, **IEEE Sensors Letters**, vol. 4, no. 8, July 2020.
  2. **D. Jyotishi** and S. Dandapat, “An ECG Biometric System Using Hierarchical LSTM With Attention Mechanism”, **IEEE Sensors Journal**, vol. 22, no. 8, pp. 6052-6061, December 2021.
  3. **D. Jyotishi** and S. Dandapat, “An Attentive Spatio-Temporal Learning-Based Network for Cardiovascular Disease Diagnosis,” **IEEE Transactions on Systems, Man, and Cybernetics: Systems**, vol. 53, no. 8, pp. 4661 - 4671, March, 2023.
  4. S. Das, **D. Jyotishi** and S. Dandapat, “Heart Valve Diseases Detection Based on Feature-Fusion and Hierarchical LSTM Network”, **IEEE Transactions on Instrumentation and Measurement**, vol. 71, Sept. 2022.
  5. S. Das, **D. Jyotishi** and S. Dandapat, “Automated Detection of Heart Valve Diseases Using Stationary Wavelet Transform and Attention-Based Hierarchical LSTM Network”, **IEEE Transactions on Instrumentation and Measurement**, vol. 72, April, 2023.

- Manuscripts Communicated

  1. **D. Jyotishi** and S. Dandapat, “Learning Enhanced Morphological Representation and Multiscale Temporal Dynamics for ECG Based Biometric Application”.
  2. **D. Jyotishi** and S. Dandapat, “Identity ECG (iECG) Based Feature Conditioning for Person Adaptive Automated CVD Diagnosis”.

Conference and Workshop Publications

- Published Paper and Accepted Publication:

1. D. Jyotishi and S. Dandapat, “An Attention Based Hierarchical LSTM Model for Detection of Myocardial Infarction,” *in IEEE 17th India Council International Conference (INDICON),* New Delhi, India, 2020 .
2. D. Jyotishi and S. Dandapat, “Person Identification using Spatial Variation of Cardiac Signal,” *in IEEE Applied Signal Processing Conference (ASPCON),* Kolkata, India, 2020.
3. D. Jyotishi and S. Dandapat, “ A Multi-Scale Residual Neural Network for ECG Based Person Identification,” *in IEEE 19th India Council International Conference (INDICON)*, Kochi, India,2022.