[UCI]I804:24011-200000342414

2020년 8월

박사학위 논문

Data Independent Acquisition based Multi-resolution Deep Networks for Biometric ECG Authentication

조선대학교 대학원

컴퓨터공학과

Htet Myet Lynn

Data Independent Acquisition based Multi-resolution Deep Networks for Biometric ECG Authentication

생체인식 ECG인증을 위한 데이터 독립 수집 기반 다중 해상도 딥 네트워크

2020년 8월 28일

조선대학교 대학원

컴퓨터공학과

Htet Myet Lynn

Data Independent Acquisition based Multi-resolution Deep Networks for Biometric ECG Authentication

지도교수 김 판 구

이 논문을 공학박사학위신청 논문으로 제출함.

2020년 5월

조선대학교 대학원

컴퓨터공학과

Htet Myet Lynn

Htet Myet Lynn의 박사학위논문을 인준함

정 일 용	조선대학교	교 수	정일용 (인)
반 성 범	조선대학교	교 수	반성범 (인)
최 준 호	조선대학교	교 수	최준호 (인)
황 명 권	KISTI	교 수	황명권 (인)
김 판 구	조선대학교	교 수	김판구 (인)

2020년 7월

조선대학교 대학원

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Data Independent Acquisition based Multi-resolution Deep Networks for Biometric ECG Authentication

Htet Myet Lynn
Advisor : Prof. Pankoo Kim, Ph.D
Department of Computer Engineering
Graduate School of Chosun University

Due to the unique intrinsic characteristics of electrocardiogram (ECG) which satisfied all categorical biometrics traits, ECG signal is gaining wide attention to be used as a reliable form of biometric for personal identity authentication applications. ECG signal was strongly approved for being discriminative to identify individuals from a relatively large population, and it also allows verification, authentication and identification.

In this dissertation, the study of both fiducial and non-fiducial based approaches for ECG biometric authentication are examined, and proposed multiple excessive techniques in order to carry out the comparative experiments to evaluate the best possible approach for all classification tasks among them. Fiducial approach relied on the P-QRS-T complex which are the set of points represent the ventricular activation of a heartbeat in the ECG signal. Those points are annotated by peak detection process to extract latency and amplitude features of a particular signal. Upon the similarity of those stored signals in database and features from compared signal are then be measured to identify a designated person from the database. However, non-fiducial methods are designed to extract the discriminative information of the signal without annotating fiducial points or extracting latency and amplitude features to

perform classification. It still needs to perform peak detection to indicate a heartbeat. As a result of recent studies usually relying on heartbeat segmentation, QRS detection is required, and the process can be confused for ECG signals where QRS complex is absent. Thus, many studies only conduct biometric authentication task on ECG signal with QRS complex and suffer from some similar limitations. To overcome this issue, we proposed a data independant acquisition method to enable highly generalizable signal processing and feature learning process by enhancing random segmentation to avoid complicated fiducial feature extraction. Then, to enrich the data representation, auto-correlation process is performed to eliminate the phase difference due to the random segmentation. Afterward, Recurrent Neural Network(RNN) with Long-short-term memory (LSTM) deep networks in bi-directional manner(BLSTM) is applied to automatically learn the features from the signal, and perform authentication task.

In addition, to investigate the effectiveness of the proposed method, we designed several deep networks models such as 1D-Convolutional Neural Network(1D-CNN), RNN-LSTM, and RNN with Gated Recurrent Units (RNN-GRU) networks for experiments over six electrocardiogram datasets with diverse behaviors with different sensor placement methods along with recent similar RNN based methods. The experimental results suggested that our proposed method of data independant approach with BLSTM network achieves a relatively higher classification accuracy in every different dataset among the compared techniques, significantly scored higher accuracy rate in experiments using ECG signal without QRS complex. The results also showed that data dependent methods can only perform well for specified data type and amendments in data variations, while the proposal can also be considered to generalize to other quasi-periodical biometric signal based classification tasks for our future study.

요약

생체인식 ECG인증을 위한 데이터 독립 수집 기반 다중 해상도 딥 네트워크

택 미얏 린
지도교수 : 김판구
컴퓨터공학과
조선대학교 IT 융합대학

모든 범주형 생체 특성을 만족하는 심전도(ECG)는 고유의 본질적인 특징으로 인해 심전도 신호는 개인 신원 인증 어플리케이션에서 신뢰할 수 있는 생체 인식 형태로 사용되는데 큰 관심을 받고 있다. ECG 신호는 비교적 많은 모집단에서 개인을 식별할 수 있는 차별성을 가지고 있으며, 검증, 인증, 식별도 가능하다는 점에서 강력하게 승인되었다.

본 논문에서는 심전도 생체인증을 위한 기준점과 비 기준점 기반 접근법의 연구를 검토하고, 제안한 방법과 비교실험을 통해 그 중에서 분류 과제에 가장 적합한 접근 방식을 평가했다. 기준점 기반 방식은 포인트 집합인 P-QRS-T 파형에 의존하며, 이는 심전도 신호에서 심장 박동의 심실 활성화를 나타낸다. 이 점들은 peak 검출 프로세스에 의해 주석을 생성하여 특정 신호의 지연 시간과 진폭의 특성을 추출한다. 데이터베이스에 저장된 신호와 비교 신호의 특징이 유사할 때 데이터베이스에서 지정된 사람을 식별하기 위해 측정한다. 그러나 비 기준점 기반 방식은 기준점에 주석을 생성하거나 지연 시간과 진폭의 특성을 추출하지 않으며, 신호의 차별적 정보를 추출하여 분류를 수행하도록 설계되어 있다. 그러나, 심장 박동을 나타내기 위해서는 여전히 peak 검출을 수행해야 한다. 일반적으로 심장 박동 분할에 의존하는 최근 연구 결과는 QRS 파형 검출이 필요하며, QRS 파형이 없는 심전도 신호에 대한 방법은 어려울 수 있다. 따라서 대부분의 연구들은 QRS 파형을 가진 신호에 대해서만 생체 인증 작업을 수행하였으며, 이와 같은 제한사항에 대한 어려움을 해결하지 못하고 있다. 이 문제를 극복하기 위해, 복잡한 기준점 특

징 추출을 피하기 위해 무작위 분할을 강화하여 고도로 일반화할 수 있는 신호 처리와 형상 학습 과정을 가능하게 하는 데이터 독립형 획득 기반 방법을 제안했다. 그다음 데이터 표현을 풍부하게 하기 위해 무작위 분할로 인한 위상 차이를 완화하기 위해 auto-correlation 프로세스를 수행한다. 그 다음 단계로, 양방향 LSTM(Long-short-term memory)으로 딥 네트워크를 구성하는 순환신경망(RNN)을 적용하여 신호로부터 형상을 자동으로 학습하고 인증 작업을 수행한다.

또는 제안된 방법의 유효성을 조사하기 위해 다양한 센서와 함께 6개의 심전도 데이터셋에 대한 실험을 위해 1차원 콘볼루션 신경망(1D-CNN), RNN-LSTM, RNN과 Gated Recurrent Units(RNN-GRU) 네트워크 등 여러 개의 심층 네트워크 모델을 설계했다. RNN과 유사한 방법을 포함한 실험 결과로 모든 데이터 집합에서 BLSTM 네트워크를 사용한 데이터 독립적 접근 방식이 비교 기법들 중 상대적으로 더 높은 분류 정확도를 달성했으며, QRS 파형을 가지지 않은 ECG 신호를 사용한 실험에서도 정확도가 크게 향상됨을 보였다. 이 결과를 통해 데이터 의존적 방법이 특정 데이터의 유형과 변경된 데이터의 수정에만 잘 수행되었으나 본 연구에서 제안한 방법을 통해 ECG 외의 생체 신호에 대한 분류를 위한 향후 연구에서도 활용할 수 있다는 것을 보였다.

Ⅰ. INTRODUCTION

The first chapter includes the interpretation of the motivation of this thesis, and it provides the aim of the thesis throughly. Section B gives a brief outlines of the following chapters.

A. Motivation

Nowadays, we encounter an emerging digitization of most areas of our everyday lives. Day-to-day we make use of online applications and services such as mobile banking, social-networking, online stock exchange and trading or email services which we do not hesitate to keep our personal confidential information on our devices or respective client severs. The digital era, sadly, has also encountered a series of new attacks and exploits, as well as unauthorized access to our sensitive information and devices by malicious virus or host. It is incredible to witness that large populations of users are still relying on numerous types or particular sets of passwords which have been used for authorized access since the earliest era of computing.

Recent years, there has been a shift of attention towards the area of biometric security systems. Such security applications support identifying an individual using their biological distinct characteristics instead of set of numerical or alphabetical passwords. The most widespread techniques imply using fingerprint, iris, and face recognition approach. normally can be found in smart devices.

In regards to biometrics in mobile devices, the benefits seem black and white. There is an added degree of security in relying, at least in part, on an extremity (e.g. finger) that only you have access to at all times. For instance, with the demand for a standard password, together with personal fingerprint, the sense of security increases. Beyond smartphone security, using features like touch ID make interacting with your device much more convenient. With applications based on such security systems, instead of having to manually enter your payment information, all you need to do is swipe your finger across your device. It's simple and saves time.

Using biometric authentication for security purposes also works to better secure information, processes, and establishments. Some organizations implement biometric scanning as a modern method of "punching in" to work. This assures that all employees are honest in terms of the hours they've worked. In turn, this saves the organization money.

However, there are still difficulties and issues related to fingerprint usability and reliability. Current challenges in ECG biometric classification tasks include extracting the features from ECG signal in order to implement a model to learn the hidden patterns for accurate generalization, proving the stability of the biometric and protecting against the attacks. In this thesis, a biometric based electrocardiogram (ECG) signals for human authentication with both fiducial and non-fiducial techniques are proposed. In addition, unlike previous studies which relied on specified data for classification, a method of data independent acquisition technique is also extended to form a model which can handle various types of ECG data input without hesitation.

B. Outline

The outline of our thesis is organized as follows:

Chapter II provides brief overview of biometrics, their characteristics and applications. It also highlights the difference between authentication and identification. The chapter consists of an introduction to the detail information of physiology behind an ECG signal and how it can be considered as a reliable tool for a biometric modality.

Chapter III includes a literature review and a insight information of existing recent studies for two different approaches, namely fiducial and non-fiducial approach concerning ECG biometric classification tasks, including the explanation of corresponding adapted techniques applied for each approach.

Chapter IV defines the conceptual design, exploited heuristic and detailed description of the proposed deep network methods and construction of their architectures. It also provides the task of ECG based biometrics human identification based on RNN networks in bidirectional manner learning technique for both LSTM and GRU cell unit, which currently perform a significant performance in the field of machine learning.

Chapter V provides the experimental reports which showed the proposed models outperform the recent state-of-the-art techniques by adapting the data independant acquisition with bidirectional learning trait training process. It also suggests that it can predominantly improve the performance for generalization complex hidden patterns. The classification accuracy for all proposed methods and other recent studies also compared.

Chapter VI concludes a brief summary, and discussion of our studies whereas proving our future research directions.

II. BACKGROUND

A. Biometrics

The term 'biometrics' is used to describe measurable and distinctive characteristics that can be used to perform recognition of individuals. These characteristics are often divided into two categories: physiological and behavioural [41]. Physiological biometrics relate to human physiology; these include fingerprints, facial features, iris patterns or DNA. Behavioural biometrics are based on human behavior, such as keystroke dynamics, voice or gait.

Biometrics are becoming increasingly used in access control and user authentication. Most existing security applications require using something that you know (e.g. passwords) or something that you have (e.g. secure tokens). With biometrics, it is possible to use something that you are, which improves system usability, as users are no longer required to remember any secrets or always carry a physical token. Access control applications can also combine multiple modalities, which improves security even further.

In order for a biometric to be applicable for access control, it must have the following characteristics [41]:

B. Authentication and Identification

Biometrics can be used to achieve two important access control goals, user authentication and identification. Biometric authentication involves the user presenting an identity claim and a biometric sample. The system then decides whether this claim is valid based on the recorded biometric for this identity. For instance, presenting your passport at the border can be seen as user authentication, where you claim the identity of the person to whom the passport belongs. In contrast, user identification involves finding the closest match to presented biometrics among the stored records. In this case, there are no identity claims. For instance, identifying wanted criminals from CCTV footage is an example of user identification. Identification is further divided into closed-set and open-set. In the former case, it is assumed that there is a biometric sample of the user that is already stored in the system. Otherwise, the problem is considered open-set.

Biometric authentication or identification is often performed by template matching. More specifically, a biometric template is a recorded instance of the biometric characteristic collected during the enrolment of the subject in the system. This template is stored in a gallery of templates. A query consists of recording a sample of the biometric during the operation of the system. This query is then matched against templates specific only to the claimed identity (in authentication) or against all templates in the gallery (in identification) [38].

Throughout this report, I will focus my attention on biometric authentication, as opposed to identification. Nevertheless, the ideas presented in this report can be also readily applied when designing an identification system.

C. Physiology Of the Heart

Before discussing the utility of an electrocardiogram as a biometric, this report presents an overview of the electrical conduction system of the heart and how it relates to electrocardiograms.

The heart is the muscle that pumps blood filled with oxygen and nutrients through the blood vessels to the body tissues [43]. The heart contains four chambers: the upper two chambers (left and right atria) are entry-points into the heart, while the lower two chambers (left and right ventricles) are contraction chambers sending blood through the circulation. The cardiac cycle refers to a complete heartbeat from its generation to the beginning of the next beat, comprising several stages of filling and emptying of the chambers. The frequency of the cardiac cycle is known as the heart rate (measured in beats per minute, bpm) [20].

In order to pump blood, the heart muscle must contract, which requires an electrical impulse. This impulse comes from the sinus node (located in the right atrium), which is transmitted via specific pathways throughout the heart, enabling regular contraction and relaxation [7]. The electrical impulse generated by the heart can be detected on the surface of the body using electrodes placed on the skin, which is done during an electrocardiogram (ECG or EKG) test. An ECG trace captures the process of depolarisation and repolarisation of the heart chambers, which causes them to contract and relax. The connection between an ECG and the electrical activity of the heart can be seen from Figure 2.1.

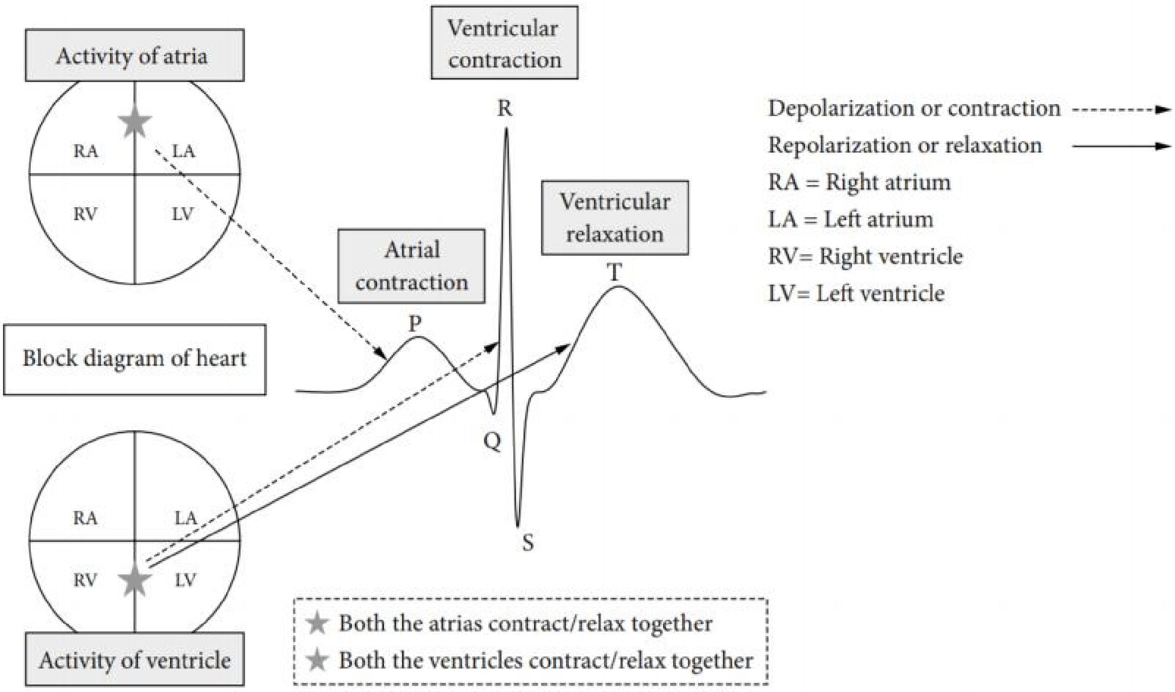

Figure 1. Generation of an ECG trace from electrical activity of the heart. [42]

ECG monitors are used to record the electrical activity of the heart using pairs of electrodes placed on the skin. Each pair of electrodes is known as a lead and provides an electrical view of the heart from a different angle. There are 12 leads that are used in cardiology, obtained from a combination of 10 electrodes. Different ECG monitors are distinguished by the number of leads that they can record [3].

An ECG trace (for some specific lead) for a single cardiac cycle consists of several parts [3]:

PR interval. The time between the beginning of the P wave and the beginning of the Q wave.

P wave. Corresponds to atrial depolarisation.

PR Segment. The time between the end of the P wave and the beginning of the Q wave.

QRS complex. Corresponds to ventricular depolarisation.

ST segment. The time between the end of the S wave and at the beginning of the T wave.

T wave. Corresponds to ventricular repolarisation.

QT interval. The time between the beginning of the QRS complex and the end of the T wave.

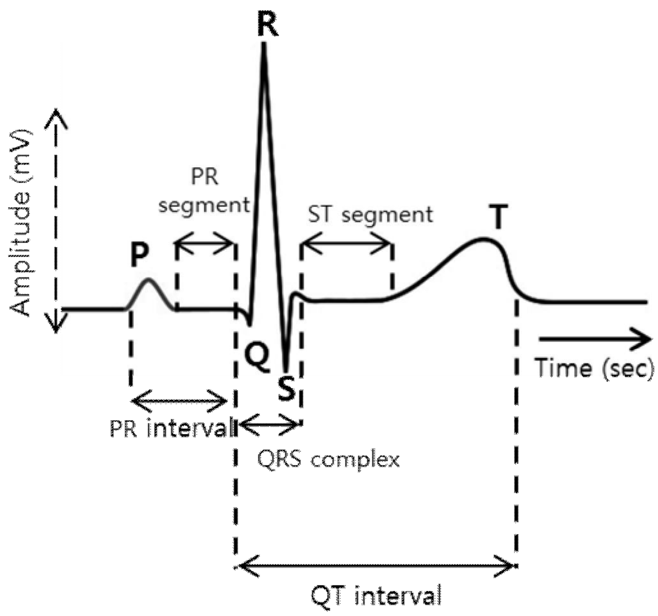

Figure 2. A single cardiac cycle of the ECG signal.

A visualisation of an ECG for a single cardiac cycle is presented in Figure 2.2. Additionally, we can also measure the RR interval, which starts at the peak of one R wave and ends at the peak of the next R wave. RR intervals can be used to compute the heart rate from a recorded ECG signal.

D. Electrocardiogram Biometric

Having discussed the physiology behind an electrocardiogram, we can now consider whether it could be used as a viable biometric. Recall that a biometric is applicable for access control if it is universal, unique and stable. Clearly, ECG is universal, as it is conditional on the electrical activity of the heart, which occurs in every living individual. Most existing work, therefore, focuses on establishing uniqueness and stability of ECG.

1. Uniqueness

Some authors claim that the composition and activity of the human heart is unique, as it inherits uniqueness from the individuality of DNA [20]. The argument that they provide is that by the “central dogma” of molecular biology, genetic information flows from the DNA to RNA (ribonucleic acid) to proteins, which are responsible for the structure and regulation of internal organs, including the heart. Using this argument, we can conclude that ECG of each individual is caused by a unique set of factors. However, the inverse, that each ECG is unique because it is produced by unique set of factors, does not necessarily follow. Therefore, uniqueness of ECG for practical applications needs to be established with empirical evidence.

Most works that explore ECG for personal identification do not assess the performance of their ECG authentication systems on very large datasets, as was done for other biometric modalities. A notable exception is a study by Carreiras et al., which focuses on the uniqueness of ECG signals [8]. The authors of the paper evaluated the performance of their biometric system on a database of ECG recordings collected from 618 subjects using a 12-lead ECG and obtained high recognition rates. The results from this work provide a positive outlook on the issue of ECG uniqueness.

2. Stability

Considerably fewer studies investigate the stability of ECG signals. While proving uniqueness can be achieved using data from a single point of time, proving stability requires data to be collected from the same individual over a sufficiently long period of time. Creating large databases of such longitudinal data is expensive and involves significant time investment, which explains the small number of studies that examine ECG stability. A study by Silva et al. collected ECG data from 63 subjects, with two data acquisition sessions separated by a 4-month interval [12]. Their results indicate that biometric authentication performs worse for longitudinal ECG data, but is still viable for real-world applications.

Additionally, we would like to consider collectability, performance, acceptability and circumvention characteristics of ECG signals. A discussion of these four characteristics is presented below.

3. Collectability

Traditional 12-lead ECG machines require 10 self-adhesive electrodes to be placed on the subject chest and limbs. While such machines do not require any effort from the subject to perform an ECG recording, they are often stationary, expensive and take time to set up. Recording ECG using medical-grade monitors is also invasive, requiring the subject to expose their chest and limbs. With the rise of personalised healthcare, however, consumer-oriented ECG monitors are becoming more widespread. These monitors are portable and can be used to record a single-lead ECG trace using electrodes that make contact with the wrists (e.g. smartwatch bands) or fingers (e.g. sensors installed on surface). While consumer-grade ECG monitors provide less data than 12-lead ECG machines, they can be used to record ECG in a non-invasive manner, applicable for biometric systems.

4. Performance

The performance of a biometric system also depends on the quality of signal preprocessing and feature extraction. Some biometrics have established methods for transforming the raw signal to features that are used for recognition of individuals. Fingerprint scanners, for instance, detect very specific fingerprint features called minutiae, which are used to establish the similarity between a biometric template in the system and a query. ECG as a biometric is much less researched and, thus, there is less consensus over which features should be used. Furthermore, some existing ECG preprocessing techniques are computationally expensive to perform, which might prevent the deployment of ECG-based biometric systems on a large scale.

5. Acceptability

With the introduction of reliable consumer-grade ECG sensors, there has been more opportunities to create ECG-based biometric systems that are non-invasive and socially accepted. Particularly appealing are “off-the-person” approaches for signal acquisition, in which biometric sensors are embedded into existing systems, such as keyboards, ATM panels and vehicle steering wheels [8]. Nevertheless, there have been no known studies that investigate the attitude of users towards using ECG as a biometric.

As with other biometric modalities, using ECG data for personal identification poses considerable security and privacy concerns. For instance, compromised ECG signal can be used to learn about certain health conditions of enrolled users. Therefore, secure storage and usage of biometrics is required to ensure that the biometric system is trusted by its users.

6. Circumvention

All biometric systems are subject to presentation attacks, which attempt to subvert the system with an artifact or contraption. Nevertheless, biometrics differ in the amount of time and resources that is required to design a suitable artefact. In order to compromise an ECG recording, the attacker has to steal the records from a medical institution or perform a social engineering attack to manipulate the victim into giving their ECG. Once that is achieved, the adversary has to digitalise the recording (if it is on paper) and fake the voltage levels at the electrodes of an ECG sensor using a device that outputs electrical waveforms (e.g. an arbitrary waveform generator). The second part of this attack was demonstrated by Eberz et al., who shows that technological barriers for the attacker are extremely low [14]. A common way to counter presentation attacks is by using liveness detection, which aims to detect whether the biometric is presented by a living individual. While some authors claim [24] that ECG offers an inherent liveness detection (being only present in a living subject), this still does not resolve the problem of presentation attacks via signal injection. More work in this area is required to establish a viable defence against ECG data compromise.

To summarise, ECG remains a strong candidate to be used as a biometric for personal recognition. Several studies have demonstrated uniqueness and stability of ECG, albeit on a small scale. The introduction of low-cost ECG sensors also provides an opportunity for system designers to embed these sensors into existing access control systems. At the same time, there is still insufficient research into extracting features from ECG signals, preventing spoofing attacks and guaranteeing that ECG-based biometric systems are accepted by the general public.

III. RELATED WORK

A. Fiducial Methods

ECG based authentication applications are commonly based on two approaches, namely fiducial and non-fiducial methods. The fiducial approach insists feature extraction process, where the points of interest within the heartbeat wave. Usually, the heartbeat wave consists of P-QRS-T complex captured as the activity of a beating human heart while recording with electrodes on human body parts. These complexes are then used to extract latency and amplitude features [4][6]. Such approaches generally rely on robust heartbeat segmentation and fiducial peak point detection. Thus, the manual feature engineering efforts are necessary in order to capture a heartbeat from original ECG signal, and it is important to be able to precisely estimate to annotate the peak locations of the P-QRS-T complex of a signal. After detection of such fiducial information, the amplitude and time-interval between corresponding points are measured. The whole process of indicating peak data points of QRS complex, and calculating time-interval features are considered as pre-processing or signal processing phase in ECG based biometric authentication applications, see Fig. 3.

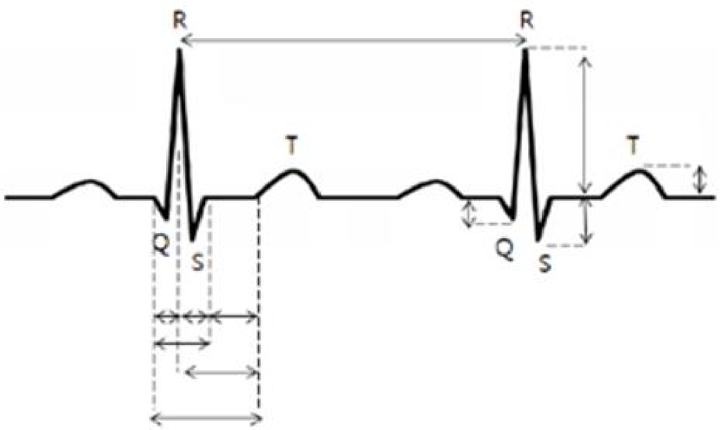

Figure 3. Fiducial points of P-QRS-T complex of ECG signal

B. Non-fiducial Methods

In the field of machine learning engaging with complex data to be generalized, deep learning methods have been successfully applied to many cases such as hand-writing recognition [10], facial recognition [7][8][9], image classification [11][12], and object recognition [14]. There are many recent studies have been proposed in the field of biometric signal based secruity systems by engaging deep learning methods [14][15][16][17][19]. Q. Zhang et al. [20] proposed a multi-resolution network based on 1D-CNN for ECG human identification applications for smart devices, and the method extended to transforming the raw input signal into multiple versions of wavelets to improve the context representations of signal, However, auto-correlation of segmented windows and transformation of wavelet are needed. X. Zhang [21] proposed models with RNN networks in various types of cell unit in hidden layers. The result suggested that the use of both LSTM and GRU gates were not significantly different for performance in terms of classification accuracy. M. Al Rahhal et al [22] proposed a method implementing stacked denoising auto-encoders with sparsity constraint, and softmax layer is applied on top of the hidden representation layer as a deep neural network. M. Zihlmann et al [23] proposed two models with based on deep neural networks, CNN and a hybrid approach of combining CNN with RNN network with LSTM cell unit. The similar approach is proposed by Warrick and Homosi [19] in the same fashion, which automatically learns the hidden characteristics of a signal and identify cardiac arrhythmias of an ECG signal with the help of CNN and LSTM techniques.

1. Convolutional Neural Networks (CNN)

Convolutional neural networks are a category of deep neural networks which have proven effective in areas such as image recognition and reated classification tasks. CNN have been successful in identifying faces, objects and traffic signs. CNN uses convolutional layers to filter input data for useful information, and a non-linear activation function applied to the results of convolutional operation. The convolution operation includes combining inputs with kernel, also known as filters to form a transformed feature map. Then fully connected layer is used after the pooling process for classification. Kernel filters complete feature extraction by sliding from top to bottom and from left to right in the original matrix. Convolutional neural network is also known as a kind of multi-layer neural networks which improve the error back propagation network. CNNs are good at classifying images, especially larger images. CNN was firstly proposed by Y. Lecun and used for handwritten character recognition [35].

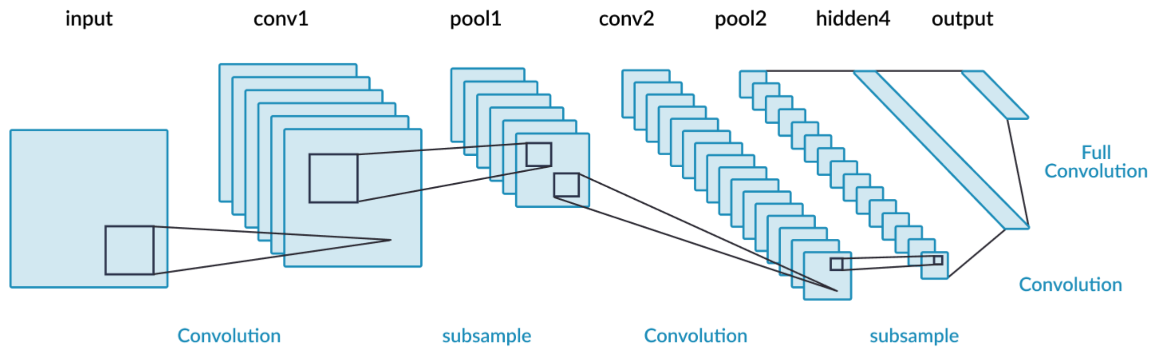

Figure 4. A sample of 7-layer CNN model classified digits for digitized pixel greyscale input images in image processing

The convolutional neural networks (CNNs) technique has two components namely feature identifier and fully connected layer. The feature identifier is carried out using convolutional layers and pooling layers, where the features are learned automatically. In ECG based classification problems, the fully connected component carries out signal classification using the features learned from the feature's identifier component.

2. Recurrent Neural Networks (RNN)

Recurent Neural Network has been a highly preferred method [28], especially for sequential data and typical RNN is illustrated as show in Fig. 5(a). Every node at a time step consists of an input from the previous node and it proceeds using a feedback loop. In RNN, each node generates a current hidden state and its output by using the given input and previous hidden state as follows:

$$h_t = f(W_h h_{t-1} + V_h x_t + b_h) \tag{1}$$

$$o_t = f(W_o h_t + b_o) \tag{2}$$

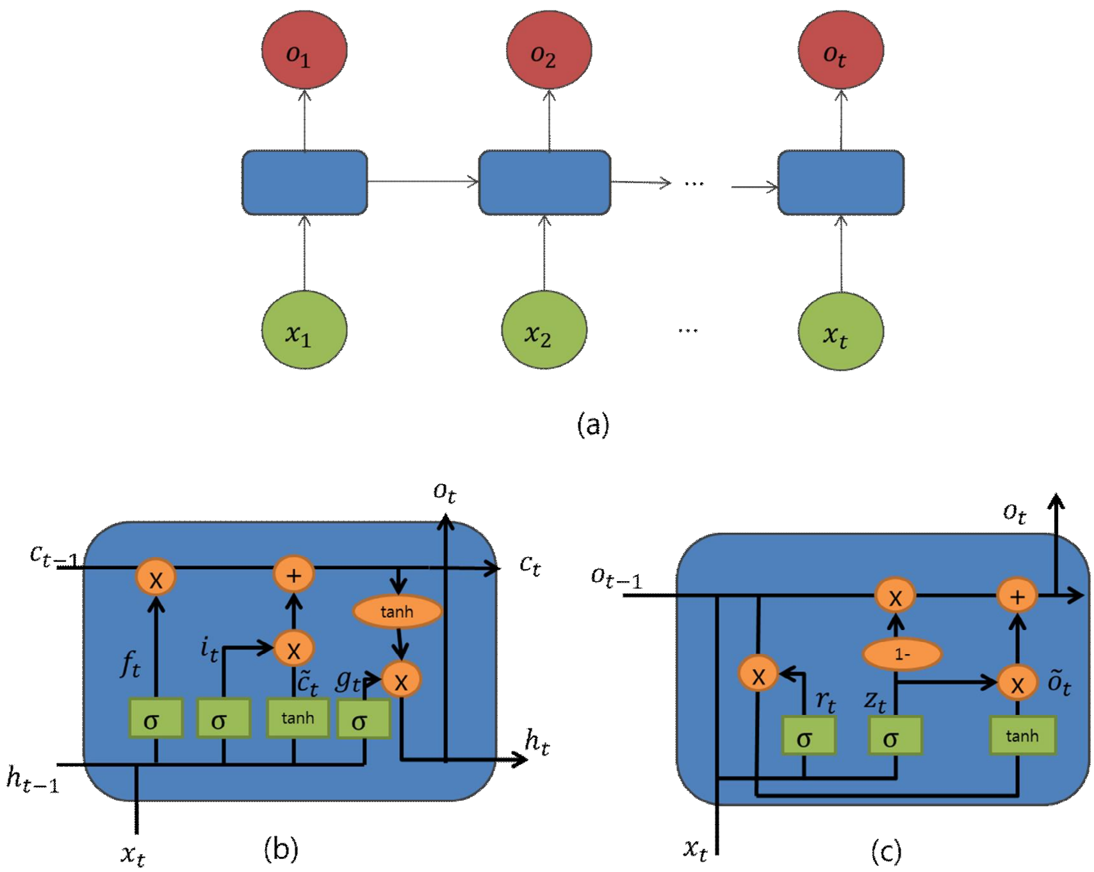

Figure 5. Recurrent Neural Networks and different cell units of its hidden layer (a) Conventional RNN Model, (b) LSTM cell unit, (c) GRU cell unit

where h_t indicates the hidden block of each time step t. W and V are the weights for the hidden layers, b denotes the bias for hidden and output states, f denotes activation function applied on each node throughout the network.

3. Long Short-Term Memory (LSTM)

Long short-term memory is a type of RNN model which designed to avoid output of a neural network for a given input from either exploding or decaying (long term dependency) as it passes through the feedback loops. Such feedback loops in RNN allow the network to be better for pattern recognition compared to other neural networks. Due to their ability to learn long term dependency, LSTMs are applicable to a number of long sequence learning problems such as language modeling and machine translation, and many other related tasks. LSTM models are designed by applying memory cells with several gates in a hidden layer. The hidden layer blocks with LSTM cell unit, and three functions of gate controllers are formulated as follows:

- Forget gate f_t decides which part of long-term state c_t should be omitted.
- Input gate i_t controls which part of $\tilde{c}_t$ should be added to long-term state c_t.
- Output gate g_t determines which part of c_t should be read and outputs to h_t and o_t.

The following equations calculate the long-term and short-term states of the cell and the output of each layer in time step.

$$f_t = \sigma(W_{xf}^T . x_t + W_{hf}^T . h_{t-1} + b_f) \quad (3)$$

$$i_t = \sigma(W_{x,i}^T . x_t + W_{hi}^T . h_{t-1} + b_i) \quad (4)$$

$$o_t = \sigma(W_{x,o}^T . x_t + W_{ho}^T . h_{t-1} + b_o) \quad (5)$$

$$g_t = \tanh(W_{xg}^T.x_t + W_{hg}^T.h_{t-1} + b_g) \tag{6}$$

$$c_t = f_t \otimes c_{t-1} + i_i \otimes \tilde{c}_t \tag{7}$$

$$o_t, h_t = g_t \otimes \tanh(c_t) \tag{8}$$

where $W_{xf}, W_{x,i}, W_{xo}, W_{xg}$ denote the weight parameters for the connected input vector, $W_{hf}, W_{hi,} W_{ho,} W_{hg}$ denote the weight parameters of the short-term state of the previous time step, and $b_f, b_{i,} b_o, b_g$ are bias.

4. Gated Recurrent Unit (GRU)

Generally, both LSTM and GRU cell unit are applied with the intuition to avoid the vanishing gradient problem in deep neural networks. While LSTM has a complex structure compared to GRU which is much simpler. The two vectors in LSTM cell are concatenated into a single vector o_t. One gate unit controls both forget and input gates. GRU is modified with a update gate to decide whether to pass previous hidden layer output to next cell or not. Forget gate is implemented for additional mathematical operation with a new set of weights. Intuitively, the reset gate decides to combine the new input with the previous memory. The update gate determines which part of the previous memory information should be passed on to the network in order to calculate the new state. The insight information and structure can be referred to Figure 5(c), and formulations for each gate and their outputs are calculated as follows:

$$r_t = \sigma(W_{xr}^t.x_t + W_{or}^t.o_{t-1} + b_r) \tag{9}$$

$$z_t = \sigma(W_{xz}^t.x_t + W_{oz}^t.o_{t-1} + b_z) \tag{10}$$

$$\tilde{o} = \tanh(W_{x\tilde{o}}^t.x_t + W_{o\tilde{o}}^t.(r_t \otimes o_{t-1}) + b_{\tilde{o}}) \tag{11}$$

$$o_t = z_t \otimes o_{t-1} + (1 - z_t) \otimes \tilde{o}_t \tag{12}$$

where $W_{xr}, W_{xz}, W_{x\tilde{o}}$ denote the weight matrices for the corresponding connected input vector, $W_{or}, W_{oz}, W_{o\tilde{o}}$ represent the weight matrices of the previous time step, and $b_r, b_z, b_{\tilde{o}}$ are bias.

IV. METHODOLOGY

A. Data Argumentation and Signal Processing

In this research, ECG-ID (ECGID), MIT-BIH Arrhythmia Database (MIT-BIH ECG), STAFF-III, and LT-AF [26][27][29][30] for ECG signal with QRS complex dataset, and AFDB[31], and AHA dataset[32] for ECG signal without QRS complex from PhysioNet, have been performed separately for all candidate models. The signal processing phase can also be determined as data pre-processing in ECG authentication applications. It mainly consists of three core operations, i.e., detrending, noisy removal (filtering), and R-peak detection which is the procedure of annotating the index data points of corresponding R-peak complex along the signal. Then the original ECG signal is detrended in order to enable the approximation better while engaging with specified length of segments for signal analysis. Nonlinear trend in the signal are also removed by fitting a low-order polynomial to the signal and subtract it, as polynominal is set to order 6 [25]. After that, Butterworth bandpass filter in the range of 5Hz and 15Hz is applied to get rid of the baseline wander. Baseline wanders are low frequency noise occurred in data acquisition in singal processing usually due to the perspiration that affects electrode impedance, respiration, body movements, for example finger movements on the electrode. The detrended signal and filtering process of original signal can be found in Fig.6. The result of filtering process are then normalized in the range of 0 and 1 by subtracting from mean value to balance the contribution for training phase using (13),(14) x and $\tilde{x}$ denote the raw ECG signal and resulted signal, respectively.

$$\tilde{x}=\frac{(x-\min(x))}{(\max(x)-\min(x))} \tag{13}$$

$$\tilde{x} = \tilde{x} - mean(\tilde{x}) \tag{14}$$

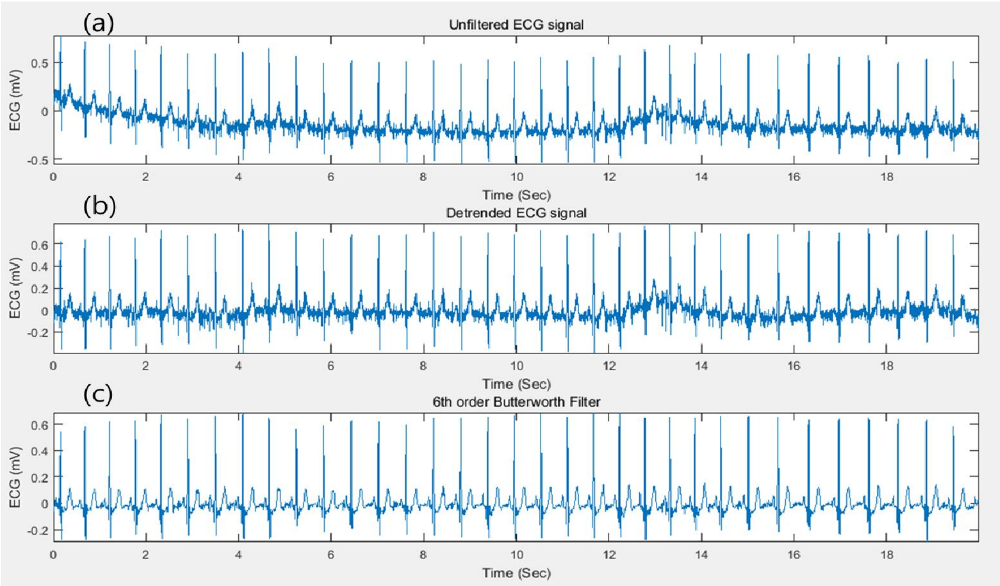

Figure 6. (a) A raw ECG signal with non-linear trend. (b) Detrended ECG signal. (c) Filtered ECG signal by applying 6th order Butterworth filter.

The ECG signal usually consists of three complexes, namely P, QRS, and T. They are accordingly determined by their corresponding complexes, also known as fiducial points which are the peak points of respective complex. Using such information as distinct features, more informative characteristics such as time domain features such as amplitudes and intervals of those complexes are generally used as features for individual signal. However, the shapes of ECG signal can be varied depend on the location of electrode on the human body parts during data acquisition. P-QRS-T complexes are not available for every version of ECG signal. Thus, non-fiducial approach with machine learning techniques are used to overcome such problems.

By introducing non-fiducial method with deep learning techniques, the fiducial extraction can be neglected in pre-processing phase. Since R peak is the most prominent peak to identify a heartbeat within the signal, R-peak detection is still needed to perform by using the Pan-Tompkins algorithms [26] for annotating the respective peak points throughout the signal. So that every heartbeat in original signal can be extracted.

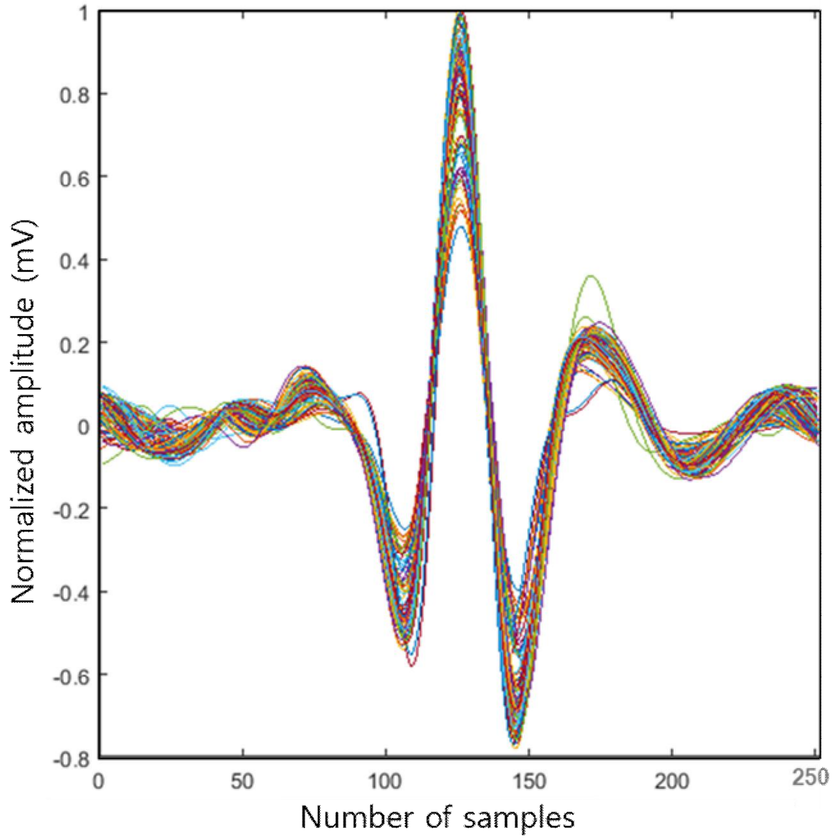

Figure 7. Multiple samples utilized for the extraction of a single heartbeat according to the respective R peak points on ECG signals of MITDB dataset.

After annotation of indices of R-peak are performed, a suitable number of samples before and after of a given R-peak point are then sliced to segment a heartbeat of a signal, which is a vector form. For our fiducial approaches, we consider 125 samples before and after of R peak point to form a heartbeat for dataset with QRS complex, and 150 samples for ECG-ID dataset. The other datasets are also conducted by same approach according to their sampling rates, a sample of a vector which interprets the heartbeat can be seen in Figure 7.

For each signal for every dataset, approximately 45 to 50 heartbeat segments are withdrawn with 251 samples, while 51 heartbeats are extracted with 301 samples from ECG-ID dataset and other datasets for heartbeat segmentation.

B. Extended Data Independent Acquisition

Unlinke the data acquisitoin technique used in previous chapter, we adapt the random segmentation without applying QRS peak detection which can be avabilable for any types of ECG signal or the signals with no QRS complex. The original signal is blindly segmented into segments with an equal length which is 2-seccond window (720 samples) to include as at least one heartbeat, since the typical rage of heart rate in a signal is from 40 to 280 beats per minute [28]. For each recording, 500 random windows are chosen, half of which are used to train and another half for testing. The auto-correlation operation is introduced to remove the phase difference due to blind segmentation.

The auto-correlation operation is applied to the segmented windows to remove the phase difference occurred by random segmentation and thus provide a shift invariant multiresolutin data representation, and it defined as:

$$Z_j^i[t] = \sum_{m=0}^{T-t-1} Y_j^i[m]\,Y_j^i[m+t],\ \forall\ t \in [0, T-1],\ \forall\ i \in [1, W],\ \forall\ j \in [1, C] \tag{15}$$

where $Z_j^i[t]$is the *t-th* sample in the *j-th* wavelet component of the *ith* ECG window after auto-correlation, $Y_j^i[m+t]$ corresponds to $Y_j^i[t]$ with a time lag of m, m is chosen from 0 to T-t-1, and T, W, and C correspond to the number of samples in an ECG window which is 720, the number of ECG window is 500, respectively.

The auto-correlation calculates the correlation of a series with its delayed copy, i.e., the similarity between series as a function of the time lag between them. Therefore, it can effectively discover repeating patterns in the quasi-periodic ECG signals even with different numbers and occurrence time of heartbeats. After removing the phase difference, the multiresolution data can now be fed to the networks for automatic feature learning and user identification purpose. Figure 8. shows similar outputs when applying auto-correlation to two wavelet domain signal segments.

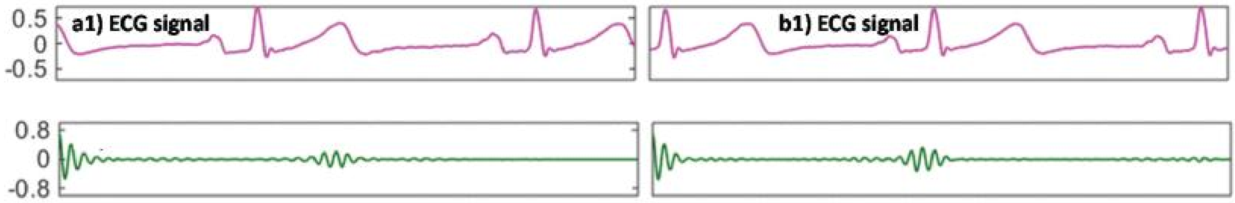

Figure 8. Multiresolution representation of two signal segments in an ECG signal

C. Models Overview

In this section, we will cover the several proposed techniques based on deep neural networks for both fiducial and non-fiducial approach for ECG authentication problem. The first model is designed based on 1-D Convolutional Neural Network (1D-CNN), which provides to learn the hierarchical distinct features to present a new version of representation of a high level abstraction. Then such abstracted informative data are fed into a classification layer such as fully connected layer for further authentication process. The rest of the other proposed methods based on RNN with modified cell units, LSTM and GRU, are also collectively proposed and investigated. Also including changes in their hidden states while the training procedure, by deploying in bidirectional manner. For conventional RNN models, the hidden state of a given time step is calculated in linear combination of the previous hidden state, and the current input. Although GRU and LSTM networks share the similar structure of network, the update gate of the hidden state is more complex in both approaches. Figure 5 illustrates the proposed RNN based models with different cell unit applied for experiments in this study.

1. Proposed 1-D CNN Model

Convolutional Neural Networks (CNNs) are neural networks built for primarily classify images, cluster images by similarity, and perform object recognition, developed in the 1980s. CNN is designed to train robustly in terms of the stochastic gradient descent algorithm for each layer. Moreover, CNNs have been commonly used for feature learning and classification problems. In this thesis, a deep 1-D CNN is designed to perform the ECG classification for fiducial approach. The detailed network architecture of proposed CNN model is implemented, and respective parameters of its network used in this study are declared according to Table 1.

Table 1. Values of proposed 1-D CNN model parameters

Layer#	1	2	3	4
Kernel size	5	2	5	2
Stride	2	2	2	2
Padding	2	0	2	0
Input size	750	375	187	94
Output size	375	187	94	47

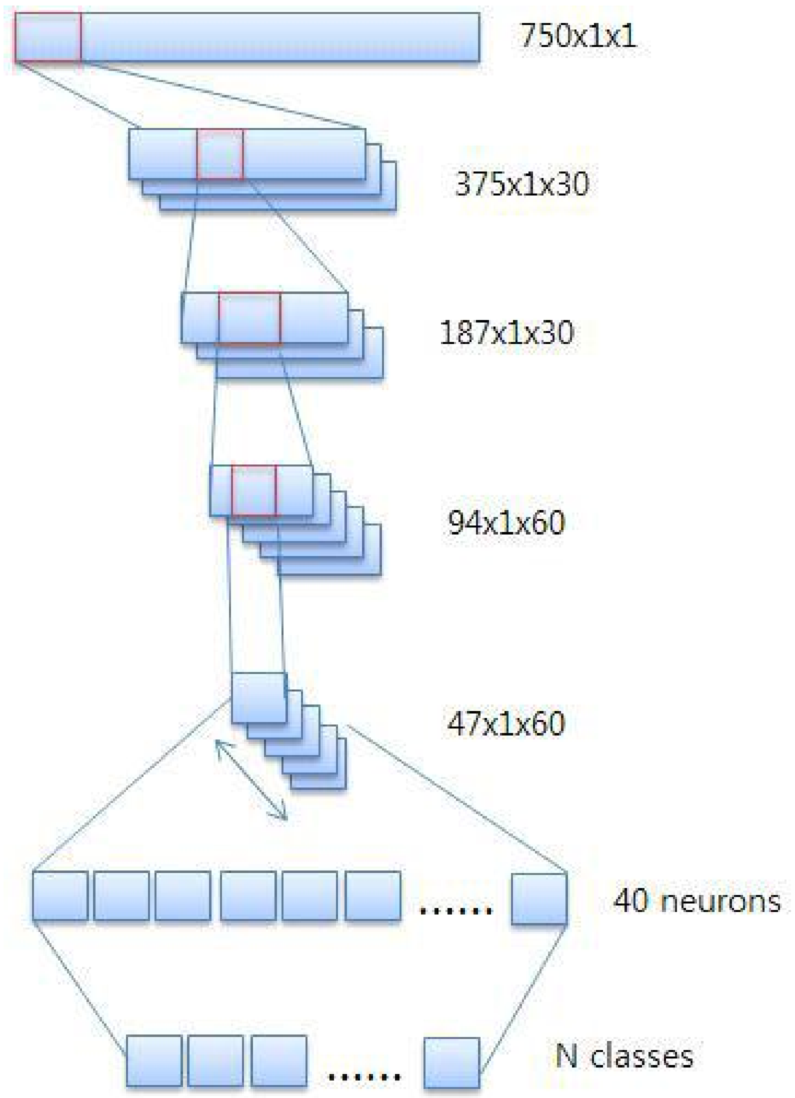

Figure 9. Proposed 1D-CNN Network architecture

There are four hidden layers in the model are used for feature learning followed by a fully-connected layer with 40 neurons. Then decision making classification layer with sigmoid function is applied to produce the appropriate categorical distribution for each class, see in Figure 9. The intuition of the proposed CNN model is that to allow a function which differentiate the patterns and distinct characteristics of all classes based on their respective input signal. Generally, the ground truth result is indicated in a one-hot distribution vector, while the input is discrete sequential samples $(x_1, \ldots x_t)$, where data point x_t is a vectorized representation of individual sample at time t. The signals are segmented into specified windows with definite length followed by the procedure presented in first section of section IV. Each window captures at least one or more heartbeat waveform in the original signal. The parameter values such as filter size, stride and padding values are set according to Table 1 throughout the network layers from first layer to the last layer. Convolutional operations with non-linear activation functions are applied between each layer. In the first and second layer, 30 filters are conducted. Finally, softmax function is used in the last layer to produce the distribution of the corresponding class for decision making in form of a vector, in the range of 0 to 1. The cross-entropy loss function for the network's targets can be calculated as:

$$E = -\sum_{i=1} (\hat{y}_i \log(y_i) + (1 - \hat{y}_i) \log(1 - y_i)) \tag{16}$$

where $\hat{y}_i$ is the ground truth target vector, y_i is the output vector of our model for i class. In order to get the output in categorical distributions across all the subjects, outputs $y_{1,..,i}$ are calculated by sigmoid function to the weight sums of activation function of the previous layer.

2. Proposed RNN Architectures

To investigate the use of RNN methods for ECG classification, we proposed the models based on different types of RNN techniques, the input training data can be set by $S=(X_n, O_n), n=1,...,N,$ where sample $X_n = x_k^n, k=1,...,m$ suggests the m numbers of samples in a signal with 251-sampled segments following the procedure from Section A for MIT-BIH, while 301-sampled segmented windows are for ECG-ID dataset. $O_n = o_i^n$, where i denotes the number of subjects. O_i^n denotes the corresponding ground truth for each subject of *nth* input. Those ground truth values are determined as 1 for given subject's signal, and indicate as 0 for the other subjects, respectively. For a given sequence input, a classifier is trained to learn the probabilities of N classes.

(i) Bidirectional RNN with LSTM cell (BLSTM) and GRU cell (BGRU)

The first proposed model for non-fiducial approach is based on a bidirectional RNN with LSTM cell unit in the hidden state layer, and is briefly called BLSTM, as shown in Figure 5(a), associated with the cell unit. The segmented signal inputs $(x_{1,...,}x_T)$ from pre-processing section, are fed into the network for each time step $t(t=1,..., T)$ for each LSTM cell. Each cell unit in a bi-directional manner consists of a parallel of LSTM tracks, known as forward and backward sequence, to capture the context from the past and future. During the final time step, those two parallel tracks of LSTM cell unit are concatenated into one single vector. In first hidden layer, the forward cell states h_0^f, the backward cell state h_0^b are initialized with zero for all layer N. The input x_t at time *t*, and previous cell states h_{t-1} to produce the output of the corresponding layer o_t^n, at time *t* and at *nth* layer for both backward or forward tracks given its parameter θ^n can be defined as:

$$o_t^n, h_t^n = LSTM^n(h_{t-1}^n, x_t; \theta^n) \tag{17}$$

$$o_t^n, h_t^n = GRU^n(h_{t-1}^n, x_{t;}\theta^n) \tag{18}$$

where θ^n denotes the parameters *(b,U,W)* of the respective cell unit for layer *n*.

For next proposed model, see in Figure 5(c), the only difference with between BLSTM and GUR gate unit is the cell unit at the hidden layers. In addition, to address one of the most important challenges in deep neural networks, overfitting, the dropout layer is applied in each cell for all RNN based methods. Sharing the similar outputs as in BLSTM at the last layer, the outputs from both forward and backward tracks, the late-fusion for bidirectional networks is concatenated into a single vector. Then the output is followed by a softmax activation function to achieve N-dimensional output in the last layer. The overall model architecture can be observed in Figure 11. As implementing based on bi-directional manner, the forward track trains the input from left to right, while the backward track traces back the input from right to left in both BLSTM and BGRU, and can be defined as follows:

$$o_t^f, h_t^f, c_t^f = LSTM^f(c_{t-1}^f, h_{t-1}^f, x_t; W^f) \tag{19}$$

$$o_t^b, h_t^b, c_t^b = LSTM^b(c_{t-1}^b, h_{t-1}^b, x_t; W^b) \tag{20}$$

$$o_t^f, h_t^f, c_t^f = GRU^f(c_{t-1}^f, h_{t-1}^f, x_t; W^f) \tag{21}$$

$$o_t^b, h_t^b, c_t^b = GRU^b(c_{t-1}^b, h_{t-1}^b, x_t; W^b) \tag{22}$$

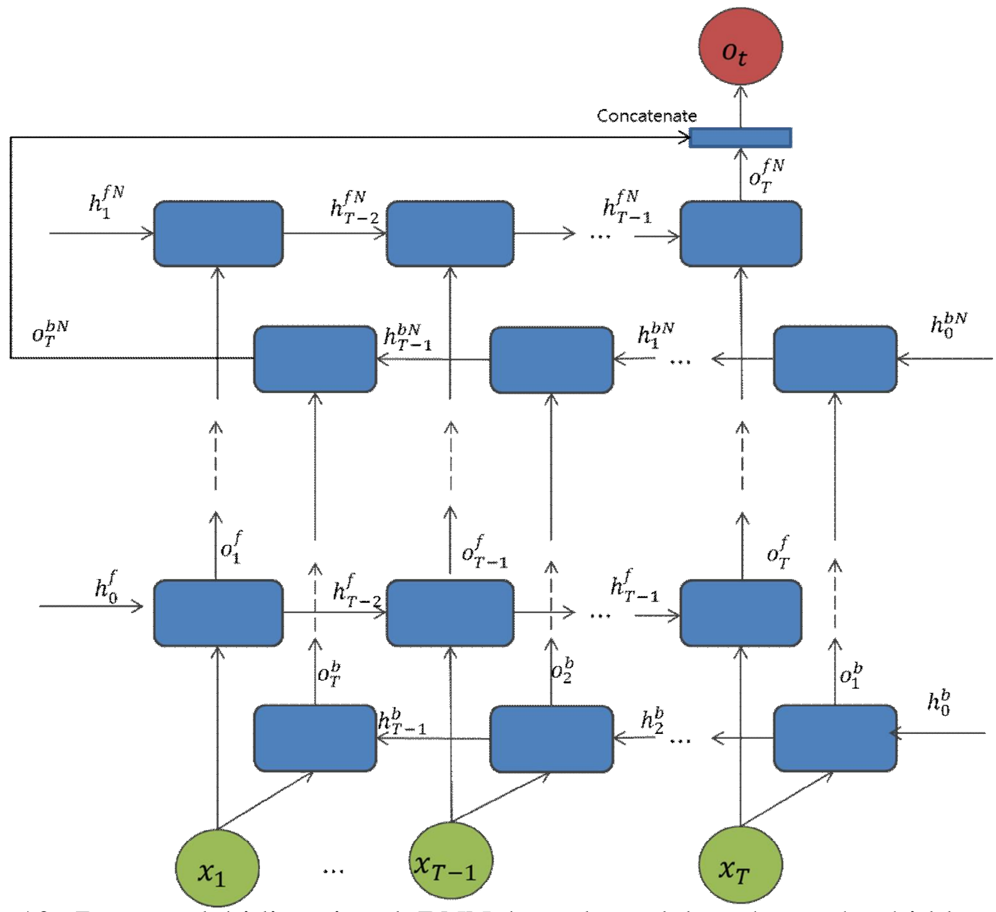

Figure 10. Proposed bidirectional RNN based models where the hidden blocks can be used as either LSTM or GRU cells

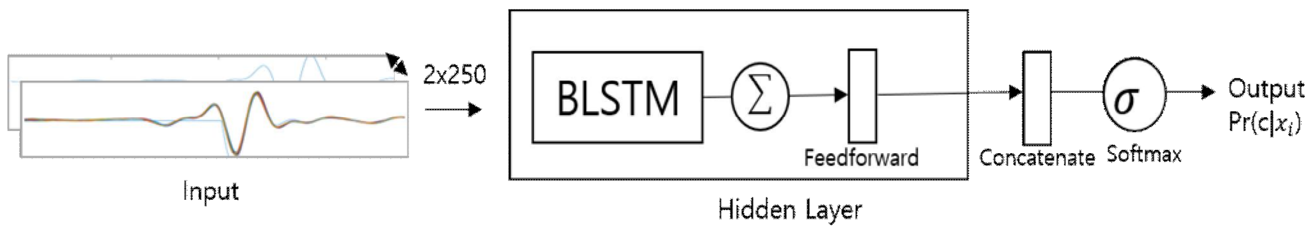

Figure 11. Overview of multi-resolution bi-directional LSTM based system architecture

V. EXPERIMENTAL RESULTS

A. Network Training

To achieve higher acceleration of the training process, usually a bottleneck while operating deep networks with many layers, our proposed models are developed in Tensorflow deep learning library, which can be executed on Graphics Processing Unit (GPU). It commonly takes at least 5 to 10 times faster than Central Processing Unit (CPU), and also can predominantly increase the training process. All our experiments are executed on GeForce GTX 1080 GPU.

During the training process, 1D-CNN based model learns hierarchical features itself by carrying out the procedure of convolution and pooling operations in accordance with the parameters provided in Table 1. A sample of the training process for 1D-CNN model is given in Fig. 12. The left block on the side suggests that the gradually increasing training and validation accuracy, while the right part corresponds to the training and validation training loss per epoch. The stochastic gradient descent (SGD) learning method is applied to increase the acceleration of the training process. That allows for passing a batch of training input data to the neural network each time. The batch size is selected as 150 for all proposed methods including RNN based networks in order to compromise two considerations, specifically a large size results in a small convergence time by reducing the variance of stochastic gradient updates, and a small size to strengthen SGD to leap out the shallow minima during the error loss function. However the network is able to learn the hidden patterns of the input signal and reached its convergence at 14 epochs, the epoch size is set as 50 to offset the under-fitting and over-fitting considerations.

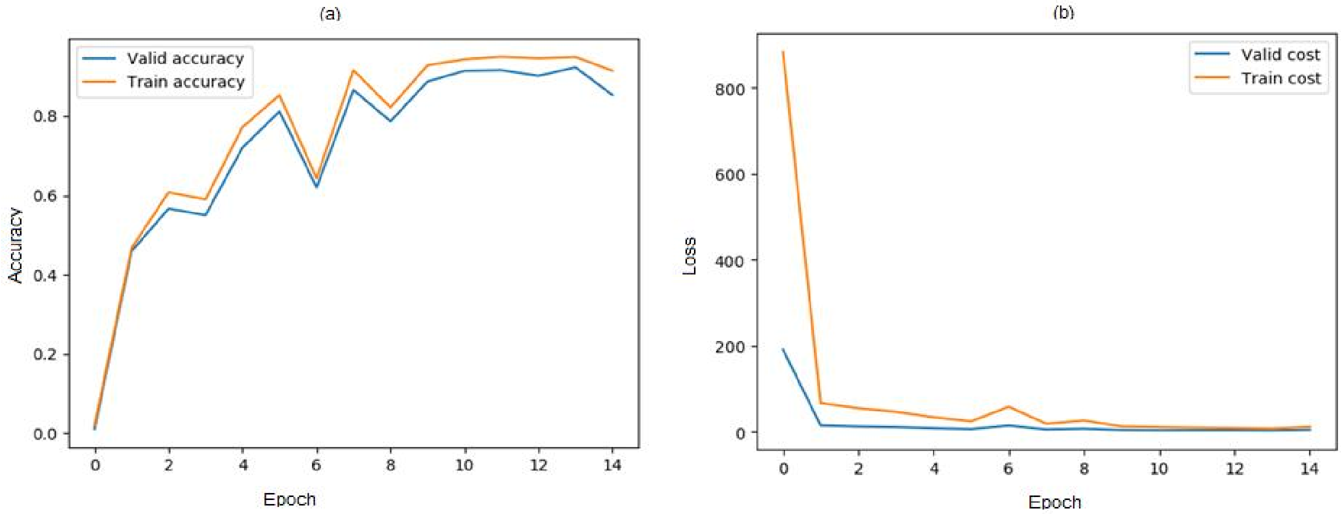

Figure 12. Accuracy and loss of the 1D-CNN model per epoch over MIT-BIH dataset: (a) Training and Validation accuracies; (b) cross-entropy loss for training and validation per epoch.

For RNN based models, the batch size is selected as of 150 since it yields better performance compared to the other schemes as mentioned above. The optimization method is applied by Adam optimizer as the learning rate is set to 0.001. Moreover, the loss functions were determined as the categorical cross-entropy method used in [47], where $\tilde{o}_l$ indicates the ground truth vector, and o_l denotes the output vector of the model for l class. For our experiments for RNN based methods, the optimal window length of segmented signal is chosen regards to the previous works and after the various length of attempts. The parameters of the proposed models are also examined with various trails of settings, and selected the optimal setting which yields better performance results. The weight parameters in the proposed models were initialized at the beginning of the training process randomly, and incrementally updated throughout the whole process. A dropout value is set as 0.2 for outputs in the first layer of the networks and the last layer inputs to avoid an over-fitting problem generally experience in learning deep neural networks. The result of cros-entropy loss, and classification accuracy rate for training and test process of the proposed BLSTM model on MIT-BIH dataset are shown in Fig. 13, respectively. The top block indicates the training cross-entropy loss as it reached to its convergence at 70 epochs when the percentage of subjects used for training is 50%.

$$E = -\sum_{i=1} \left(\tilde{o}_l \log(o_l) + (1 - \tilde{o}_l) \log(1 - o_l) \right) \tag{23}$$

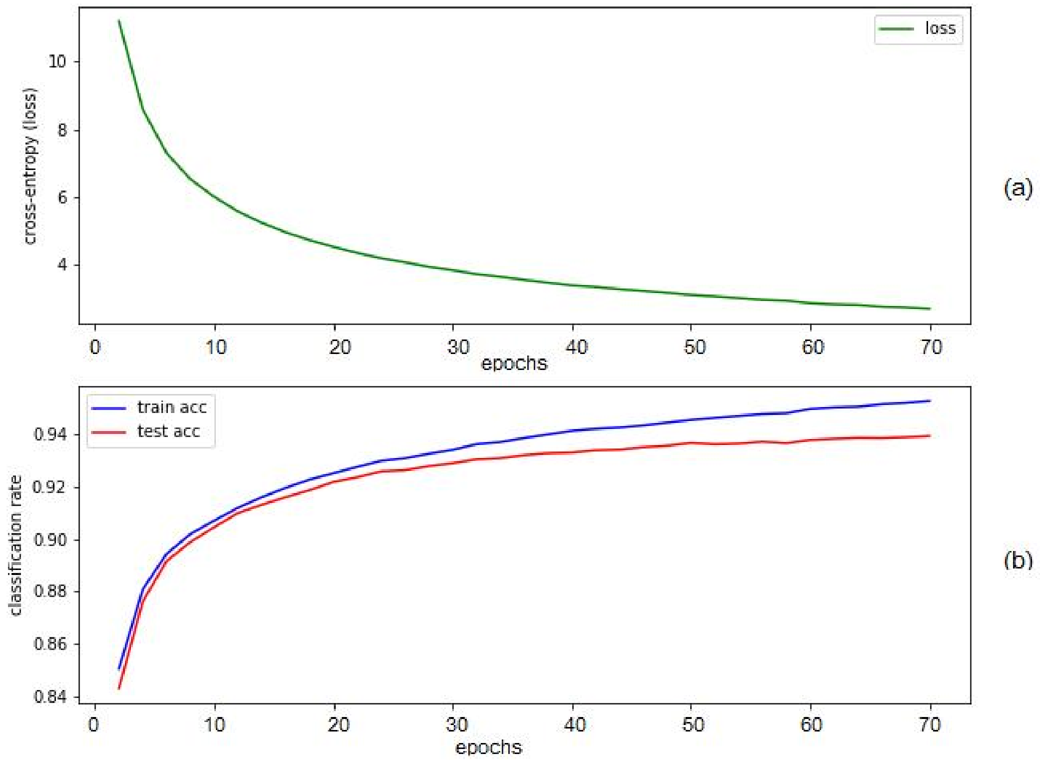

Figure 13. Accuracy and loss of the proposed BLSTM model per epoch over MIT-BIH dataset: (a) Training and Test accuracies; (b) cross-entropy loss for training per epoch.

B. System Evaluation

For our experiment, ECG-ID dataset(ECG-ID) and MIT-BIH ECG, STAFF III database (STAFF-III), and Long Term AF Database (LT-AF) data are collected from PhysioNet. ECG-ID dataset includes 310 of ECG recordings digitized at 500Hz, obtained from 90 persons (10,000 samples), while MIT-BIH ECG dataset contains 168 short recordings set to pose a variety of challenges for ECG compressors especially for compression methods. STAFF III database was acquired during 1995-96 and contains standard 12-lead ECG recordings from 104 patients. LT-AF dataset consists of 84 ECG recordings of subjects with paroxysmal or sustained atrial fibrillation (AF), and digitized at 128 Hz with durations vary which is upto 24 to 25 hours. Moreover, two more datasets, namely AFDB and AHA datasets (without QRS complex), are examined for our extended data independent acquisition based approach. To train the RNN based networks, training datatset is divided into batches of several heartbeats for each. The weights for each batch are updated upon completion of every batch. The input data is forward and backward propagated throughout the network, and error cost is calculated by back-propagating of the unfolded network in time. We adopted the method called back propagation through time (BPTT) with Adam optimization method employed for our experiment, while the learning rate is set to 0.001. The batch size is set as 150, which yields higher performance, for all methods namely traditional RNN, RNN with LSTM, and RNN with GRU. The epoch size is chosen as 150 to balance network over-fitting issues. Furthermore, a dropout of 0.4 outputs of the first layer, and the last layers inputs were utilized to overcome the overfitting problem. For the evaluation, 5066 recordings from each dataset were separated for training and test set.

To investigate the accuracy of classification rate, the proposed models were evaluated by the classification accuracy which can be determined by the confusion matrix. It is one of the most common intuitive metrics exploited for evaluating the performance and accuracy of the machine learning models commonly used for the classification problems see in Table 2.

Table 2. Confusion matrix of evaluating classification accuracy

Actual / Predicted	Positives (1)	Negatives (0)
Positives (1)	TP	FP
Negatives (0)	FP	TN

The associated terms concern with the given confusion matrix can be specified as: True positives (TP), Truen negatives (TN), False positives (FP) and True negatives (TN) respectively. When the output is classified the data point correctly as the ground truth, it is considered to be a TP. True negatives (TN) are the cases when the output class of the given data point is predicted correctly as negative for the given class. False positives (FP) are the cases when the model incorrectly predicted the corresponding class as a positive. False negatives (FN) are the cases when the ground truth should be positive as the model predicted as negative.

As a result of accuracy in classification tasks is the correct predictions made by the model over all predictions, the correct predictions known as True Positive (TP) and True Negative (TN) are divided by all predictions made by the model, calculated in (24).

$$Classification\ accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{24}$$

According to Table 3, the reported overall classification accuracy outperform all the previous studies for all datasets. Before we investigate the performance of our proposed method, we studied it against conventional RNN based methods, namely, traditional RNN, RNN with LSTM gates, and RNN with GRU gates over four datasets with QRS complex.

Table 3. Performance of classification accuracy for selected input sequence length

Type of model	Input sequence length (number of heartbeats)	Overall Accuracy
Proposed 1D-CNN	3	0.925
	9	0.941
RNN + LSTM	3	0.965
	9	0.971
RNN + GRU	3	0.952
	9	0.978
Proposed BLSTM	3	0.982
	9	0.993
Proposed BGRU	3	0.921
	9	0.983

The entropy loss plot is demonstrated in Fig. 14, and proposed multi-resolution bidirectional LSTM outperforms the others as the training process reached to 150 epochs over ECG-ID dataset. Thus, our proposed bi-directional LSTM based model is chosen for further experiments compared to our new approach with data independent acquisition based method and other recent studies.

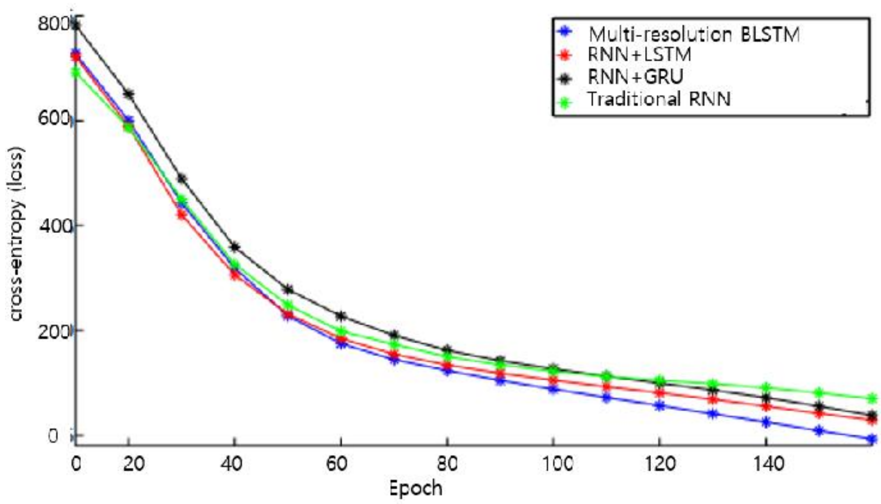

Figure 14. Comparisons of cross-entropy loss per epoch over ECG-ID dataset

To evaluate the performance results for classification, we examined the networks based on four statistical evaluation metrics, Accuracy (Acc), Sensitivity (Sen), Specificity (Spc), and Positive predictivity (Ppr).

$$Acc = \frac{TP+TN}{TP+FP+FN+TN} \tag{25}$$

$$Sen = \frac{TP}{TP+FN} \tag{26}$$

$$Spe = \frac{TN}{TN+FP} \tag{27}$$

$$Ppr = \frac{TP}{TP+FP} \tag{28}$$

The terms (TP, TN, FP, FN) in above equations denote true positive, true negative, false positive, and false negative respectively. F1 and Fowlkes-Mallows index (FM) scores are as well computed upon using Sen and Ppr as follows:

$$F1 = \frac{2}{\frac{1}{Sen}+\frac{1}{Ppr}} \tag{29}$$

$$FM = \sqrt{Sen \times Ppr} \tag{30}$$

We compared the proposed method with previous RNN based networks over two ECG datasets according to the above evaluation metrics. Our proposed multi-resolution bidirectional LSTM outperforms others in terms of **F1** and **FM** scores in ECG-ID dataset. However, the accuracy of proposed algorithm for both datasets are nearly matched with Mostayed et al. [36] which used bidirectional LSTM network for 12-lead ECG signal, note that **F1** score is a more significant metric compare to accuracy score. Q. Zhang et al. [23] used a multiresolution parallel network based on CNN deploying multiple versions of wavelets to enrich the context representation of the signal for generalization purpose. For ECG-ID dataset, the accracy for all techniques are quite similar, the proposed method surpassed the other methods in **F1** score which achieves 98.84%. For MIT-BIH ECG dataset, Fan Liu et al. [38] achieved a higher accuracy compared to other techniques, however, our proposed method significantly outperforms them, and they still need to firstly identify the heartbeat which is expensive in terms of algorithm engineering process compared with our random signal segmentation. In STAFF-III dataset, the proposed method scored a high accuracy with 97%, and achieved higer **F1** and **FM** score as well. Regarding LT-AF dataset, despite sharing the similar score with A. Mostayed et al. [36] in accuracy by achieving 99%, our proposed method significantly surpassed the others in terms of **F1** and **FM** scores with 99.5%.

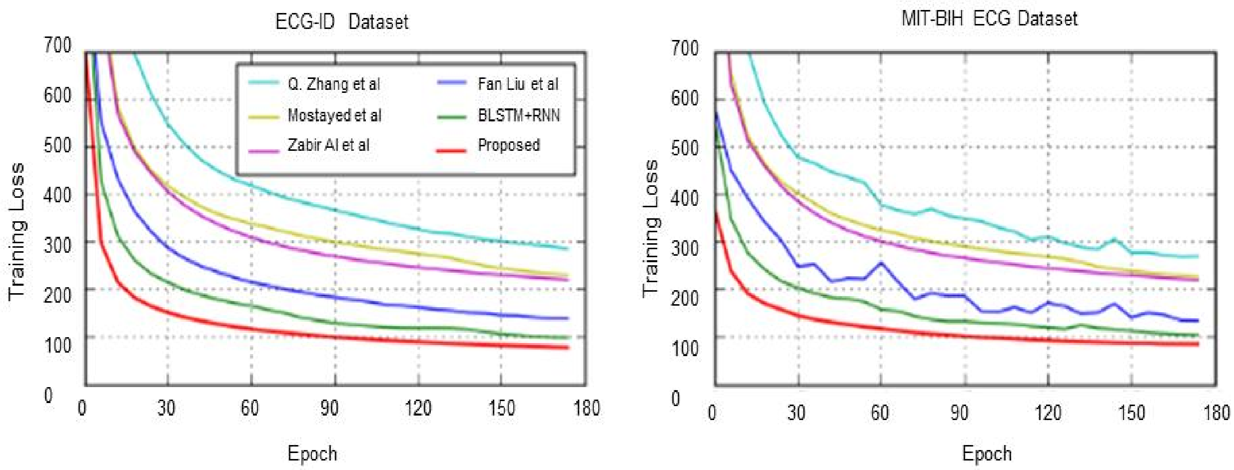

Figure 15. Comparison of cross-entropy loss per epoch over ECG-ID dataset & MIT-BIH ECG Dataset

However, for two datasets where the signals with no QRS complex, rest of the compared techniques based on R-peak detection method for heartbeat segmentation, are significantly decreased in classification accuracy. The highest score of F1 score does not come near 90.5% among all the compared techniques, while our data independent acquisition approach scored 97.94% and

97.3% for AFDB dataset and AHA dataset respectively.

Nevertheless, the proposed method provides more advantages and high enough classification capability compared to the recent studies in overall comparison in different signal versions.

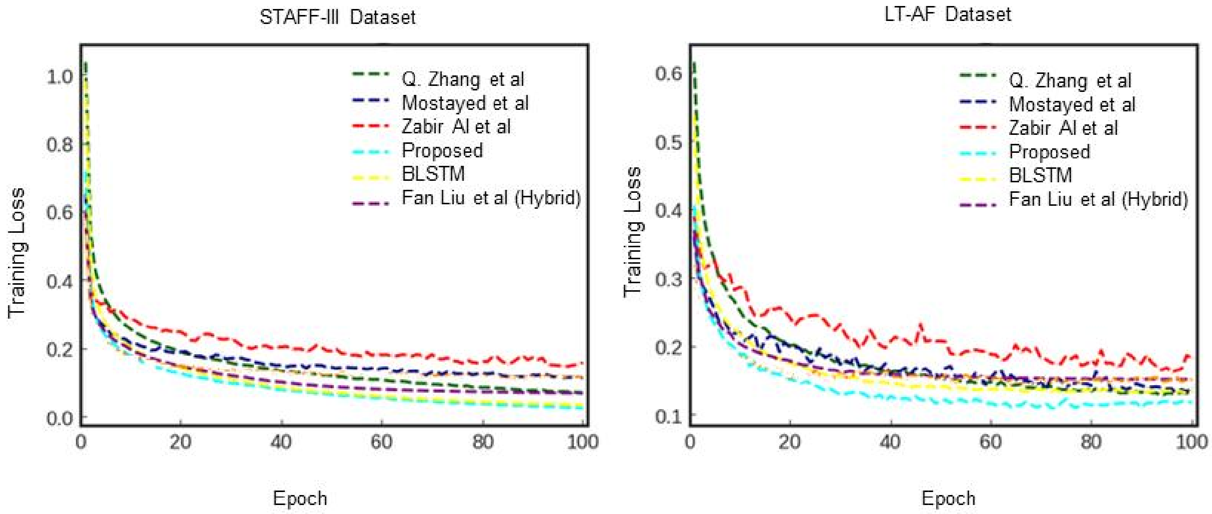

Figure 16. Comparison of cross-entropy loss per epoch over STAFF-III dataset & LT-AF Dataset

Figure. 15 and 16 show the training process for all six trials and interestingly we can find that our proposed method own best convergence speed and least training entropy loss after 120 epochs for ECG-ID dataset, and it reached to its convergent after 90 epochs for MIT-BIG ECG dataset, while the epoch rates decreased to 60 to 90 for both STAFF-III and LT-AF dataset which are considered to be presenting better signal quality with less noise and shift invariants. However, in both datasets without QRS complex, AFDB and AHA, all the other compared methods using QRS peak detection are unable to reach their convergence speed even after 170 epochs. It shows that data dependent methods can only perform well for specified data type and amendments in data variations.

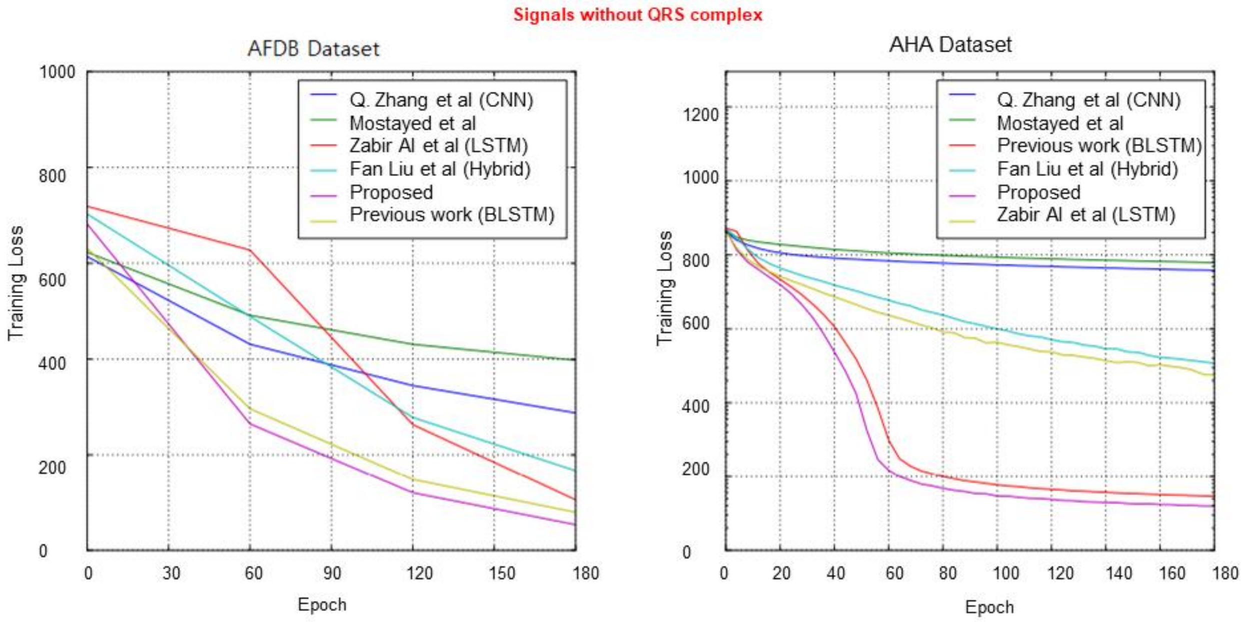

Figure 17. Comparison of cross-entropy loss per epoch over AFDB dataset & AHA Dataset

This is stable with the conceptual experiment that poor local minima are hardly an issue in deep neural networks with many layers consist of a large number of parameters. Instead, the landscape of the object function is packed with a variation of valleys which seems to commonly have local minima with similar values. Therefore, the randomness in Adam optimization based method parameter tuning process actually often results in only small fluctuations to the convergence curve in the training process.

Table 4. Comparing the proposed method with recent state-of-the-art works in classification accracy of 20 classes over four different datasets (ECG-ID, MIT-BIH, STAFF-III, LT-AF, AFDB, and EEG)

Datasets	Methods	Acc	Sen	Spe	Ppr	F1	FM
ECG-ID	Zhang.et.al. [23]	98.3	75.2	98.3	99.8	85.8	85.8
	Mostayed.et.al.[36]	98.4	93	97.5	98.2	95.5	95.5
	Zabir Al et al. [37]	90.3	94.2	95.6	93.1	93.6	93.6
	Fan Liu et al. [38]	98.3	95.7	98.2	99.2	97.4	97.43
	Previous study	98.1	96.1	98.6	98.4	97.1	97.1
	Proposed	99.3	98.3	99.2	99.4	98.84	98.84
MIT-BIH ECG	Zhang.et.al. [23]	98.6	95.2	97.3	89.5	92.2	92.2
	Mostayed.et.al.[36]	99.4	95.8	99.7	97.8	96.8	96.8
	Zabir Al et al. [37]	80.1	82.8	89.1	84.4	83.59	83.59
	Fan Liu et al. [38]	99.3	99.6	98.1	98.3	98.94	98.94
	Previous study	99.2	93	99.8	98.6	95.5	95.5
	Proposed	99.5	99.2	98.8	99.2	99.2	99.2
STAFF-III	Zhang.et.al. [23]	98.1	86.6	99.3	96.2	91.2	91.2
	Mostayed.et.al.[36]	98.7	91.3	97.4	97.8	94.4	94.4
	Zabir Al et al. [37]	89.4	88.7	92.3	89.6	89.14	89.14
	Fan Liu et al. [38]	98.6	94.6	99.2	98.5	96.5	96
	Previous study	99.1	93	98.8	98.3	95.5	95.5
	Proposed	99.3	95.5	97.9	99.2	97.31	97.07
LT-AF	Zhang.et.al.[23]	97.6	95.8	95.3	96.4	96	96
	Mostayed.et.al.[36]	99.1	99.4	98.7	98.5	98.6	98.6
	Zabir Al et al. [37]	89.4	88.4	90.2	93.2	90.7	90.76
	Fan Liu et al. [38]	99.4	99.2	98.4	97.6	98.39	98.39
	Previous study	99.5	99.7	98.8	99.2	99.4	99.4
	Proposed	99.2	99.6	98.2	99.5	99.5	99.5
AFDB	Zhang.et.al.[23]	89.6	91.3	90.5	88.7	90	90
	Mostayed.et.al.[36]	83.1	88.4	89.3	88.3	88.34	88.34
	Zabir Al et al. [37]	79.1	81.2	88.4	86.5	83.8	83.8
	Fan Liu et al. [38]	90.2	89.8	92.5	90.3	90.04	90.04
	Previous study	89.3	93.4	90.3	91.2	92.2	92.2
	Proposed	98.5	97.3	99.1	98.6	97.94	97.94
EEG	Zhang.et.al.[23]	84.3	81.5	83.6	88.4	84.87	84.87
	Mostayed.et.al.[36]	85.4	86.4	88.3	83.4	84.88	84.88
	Zabir Al et al. [37]	76.5	81.2	83.2	88.2	84.62	84.62
	Fan Liu et al. [38]	87.5	88.3	89.2	83.6	85.91	85.91
	Previous study	88.1	87.5	88.4	89.5	88.49	88.49
	Proposed	97.3	96.4	97.2	98.4	97.39	97.39

VI. CONCLUSION AND FUTURE WORK

We proposed a multi-resolution bidirectional LSTM network with wavelet compression technique for input data in ECG based biometric identification. By applying random segmentation with auto-correlation approach for independant data acquisition, it is enriched the time-frequency representation of original signal and strengthen the classification accuracy for data variations, the performance of learning procedure were improved compared to other RNN based methods and hybrid method based approaches according to the experimented results. The experimental outcomes showed that the proposed algorithm surpassed most of RNN based networks by adapting the bidirectional learning method, and considerably improved the classification performance with benefit of wavelet compression technique for more contextualized distinct features. On the other hand, we will further consider more data representation techniques and deep learning methods to explore a better feature learning capability. The proposed method of BLSTM-RNN model can also be considered for generalising to other periodical waveform of biometric signal-based user authentication applications, such as photoplethysmogram (PPG), ballistocardiograph (BCG) and body movements.

REFERENCES

[1] A. Ali. Salah. Machine learning for biometrics. In Handbook of ''Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques'', chapter 26, IGI Global, Oxford, pp. 539 - 560. 2010.

[2] N. Srivastava, G. E. Hinton, Al. Krizhevsky, I. Sutskever, and R. Salakhutdinov. ''Dropout: a simple way to prevent neural networks from overfitting''. JMLR, 2014.

[3] R. K. Srivastava, K. Greff, and J. Schmidhuber. ''Highway networks.'' ¨ arXiv preprint arXiv:1505.00387, 2015.

[4] J. Pan and W. J. Tompkins, ``A Real-Time QRS Detection Algorithm,'' IEEE Transactions on Biomedical Engineering, vol. BME-32, no. 3, pp. 230–236, 1985.

[5] Y. Bengio, ''Learning deep architectures for AI,'' Foundations and Trends in Machine Learning, 2009, vol. 2, pp. 1–127.

[6] P. Raman and S.M. Ghosh, ''Classification of Heart Diseases based on ECG analysis using FCM and SVM Methods,'' International Journal of Engineering Science, vol. 6, pp. 6739–6744, 2016.

[7] S. N. Yu and K. T. Chou, ``Integration of independent component analysis and neural networks for ECG beat classification,'' Expert Syst. Appl., vol. 34, pp. 2841–2846, 2008.

[8] O. T. Inan, L. Giovangrandi, and G. T. Kovacs, ``Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features,''IEEE Trans. Biomedical Engineering}., vol. 53, pp. 2507–2515, 2006.

[9] S. Sahoo, B. Kanungo, S. Behera, and S. Sabut, ``Multiresolution wavelet transform based feature extraction and ECG classification to detect cardiac abnormalities,'' Measurement, vol. 108, pp. 55–66, 2017.

[10] M. Thomas, M.K. Das, and S. Ari, ``Automatic ECG arrhythmia

classification using dual tree complex wavelet based features,'' AEU-International Journal of Electronics and Communications, vol. 69, pp. 715–721, 2015.
[11] R. J. Martis, U. R. Acharya, and L. C. Min, ''ECG beat classification using PCA, LDA,'' ICA and discrete wavelet transform, Biomedical Signal Processing and Control, vol. 8, pp. 437–448, 2013.
[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, ''Gradient-based learning applied to document recognition,'' Proc. IEEE 1998, pp. 2278–2324, 1998.
[13] M.Coşkun, A. Uçar, Ö. Yıldırım, and Y. Demir, ``Face recognition based on convolutional neural network,'' presented at the Int. Conf. Mod. Electr. Energy Syst. 2017. pp. 376–379, 2017.
[14] Ö. Yildirim, A. Uçar, and U. B.Baloglu, ''Recognition of Real-world Texture Images under Challenging Conditions with Deep Learning,'' ASYU, 2017.
[15] A. I. Krizhevsky, Sutskever, and G. E. Hinton, ``Imagenet classification with deep convolutional neural networks,'' Advances in Neural Information Processing Systems, pp. 1097-1105.
[16] A. Uçar, Y. Demir, and C. Güzeliş, ``Object recognition and detection with deep learning for autonomous driving applications,'' Simulation 2017, vol. 9, pp. 759–769, 2017.
[17] S. Min, B. Lee, and S. Yoon, ``Deep learning in bioinformatics,'' Briefings Bioinf. 2017., vol. 18, pp. 851–869, 2017.
[18] M.M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R.R.Yager, ``Deep learning approach for active classification of electrocardiogram signals,'' Inf. Sci., vol. 345, pp. 340–354,2016.
[19] J. H. Tan et al., ``Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals,'' Comput. Biol. Med.,, vol. 94, pp. 19–26, 2018.
[20] S. Kiranyaz, T. Ince, and M. Gabbouj, ``Real-time patient-specific ECG classification by 1-D convolutional neural networks,'' IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng., vol. 63, pp. 664–675, 2016.

[21] U. R. Acharya et al., ''A deep convolutional neural network model to classify heartbeats,''Comput. Biol. Med., vol. 89, pp. 389−396, 2017.
[22] P. Warrick, and M. N. Homsi, ''Cardiac Arrhythmia Detection from ECG Combining Convolutional and Long Short-term Memory Networks,'' doc:10.22489/CinC.2017.161-460}, 2017.
[23] 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. 11805−11816, 2017.
[24] X. Zhang, Y. Zhang, L. Zhang, H. Wang and J. Tang, ''Ballistocardiogram Based Person Identification and Authentication Using Recurrent Neural Networks,'' presented at the 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, Beijing, China, pp. 1−5, 2018.
[25] M. M. Al Rahhal et al, ''Deep learning approach for active classification of electrocardiogram signals,'' Information Sciences, vol. 345, pp. 340−354, 2016.
[26] Lugovaya T.S. Biometric human identification based on electrocardiogram. [Master's thesis] Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Saint-Petersburg, Russian Federation; June 2005.
[27] GB. Moody , RG. Mark, ''The impact of the MIT-BIH Arrhythmia Database''. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209)
[28] H. Tanaka, K. D. Monahan and D. R. Seals, "Age-predicted maximal heart rate revisited'', J. Amer. College Cardiol., vol. 37, no. 1, pp. 153-156, 2001.
[29] J. P. Martínez , O. Pahlm, M. Ringborn, S. Warren, P. Laguna, L. Sörnmo, ''The STAFF III Database: ECGs Recorded During Acutely Induced Myocardial Ischemia''. Comput Cardiol, 44:266-133, 2017.
[30] S. Petrutiu, A. V. Sahakian, S. Swiryn, ''Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans''. Europace 9:466-470 (2007).
[31] G. B. Moody, ''Spontaneous Termination of Atrial Fibrillation: A Challenge

from PhysioNet and Computers'', Computers in Cardiology 31:101-104, 2004.

[32] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, PC, R. Mark, H. E. Stanley, ``PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals''. Circulation [Online]. vol. 101, no. 23, pp. e215-e220.

[33] M. Zihlmann, D. Perekrestenko, and M. Tschannen, ``Convolutional recurrent neural networks for electrocardiogram classification,'' CoRR, 2018.

[34] Ö. Yildirim, ``A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,'' Computers in Biology and Medicine, vol. 96, pp. 189–202, 2018.

[35] Y. L. Lecun et al., ``Gradient-Based Learning Applied to Document Recognition'', Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[36] A. Mostayed, et al., ''Classification of 12-Lead ECG Signals with bi-directional LSTM Network,'' Computer Vision and Pattern Recognition, 2018.

[37] Z. A. Nazi, A. Biswas, M. A. Rayhan and T. Azad Abir, "Classification of ECG signals by dot Residual LSTM Network with data augmentation for anomaly detection," 2019 22nd International Conference on Computer and Information Technology (ICCIT), doi: 10.1109/ICCIT48885.2019.9038287.

[38] F. Liu *et al.*, ``An Attention-based Hybrid LSTM-CNN Model for Arrhythmias Classification,'' 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8852037.

[41] P. Stavroulakis, M. Stamp. Handbook of Information and Communication Security. Springer Publishing Company, Incorporated, 1st edition, 2010.

[42] F. Sufi, I. Khalil, J. Hu. ``ECG-based authentication''. In Handbook of Information and Communication Security, pages 309 - 331. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[43] Remove Trends from Data. MathWorks, 2018. https://www.mathworks.com/help/signal/ug/remove-trends-from-data.html

[44] P. S. Hamilton and W. J. Tompkins, ``Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database,'' \emph{IEEE

Transactions on Biomedical Engineering}, vol. BME-33, no. 12, pp. 1157−1165, 1986.

[45] A. L. Goldbergeret al., ``PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,''Circulation Electronic Pages, http://circ.ahajournals.org/content/101/23/e215.full, 2000.

[46] J. L. Elman, ``Finding structure in time,'' Cognit. Sciences., vol. 14, pp. 179−211, 1990.

[47] S. Hochreiter and J. Schmidhuber, ``Long short-term memory,'' Neural Compuing, vol. 9, pp. 1735−1780, 1997.

[48] K. Cho et al., ''Learning phrase representations using RNN encoder-decoder for statistical machine translation,'' CoRR, 2014.

ACKNOWLEDGMENTS

Foremost, I pay homage to the supreme triple Gems, the Buddha, Dhamma and Sangha, and give great respects to my parents and my teachers. Achieving a doctoral thesis is not an individual experience; rather it takes place in a social context and includes several persons, whom I would like to thank sincerely.

To my family, thank you all for believing in me in all of my pursuits and encouraging me to chase my dreams.

I would like to express my deepest gratitude to my advisor, Prof. Kim Pankoo, for his excellent guidance, wisdom, patience, consistent support, providing me with a wonderful atmosphere for doing research, and offering me an opportunity to expand my knowledge under his guidance since my undergraduate years.

I would like to extend my sincere gratitude and appreciation for all the helping hands and caring friendships from my colleagues in the lab, my special thanks go to Hong Taekeun. I hereby make sure to thank all the individuals who probably helped me more than they think they did throughout my studies, especially Lata.

I would also like to take time to express my appreciation towards my dearest friends, Shajeel and Sanjay, whom sometimes I forget to thank making the good times better and hard times easier, even though we are far apart from each other. I have been very fortunate to call them as friends, or my brothers from another mother.

Last but not least, I owe thanks to many people whom I cannot acknowledge fully here. I would like to thank all those who contributed their time, efforts and moral support directly or indirectly to bring my dissertation to final completion, and all my friends outside the workplace for their support which made my stay and studies in Gwangju more enjoyable. Everyone kept me going, and this journey would not have been possible without them.

Thank you to Aye Mya Swe, for all her love and support.