ONLINE AND CONTINUOUS ELECTROCARDIOGRAM (ECG) BIOMETRIC SYSTEM

by

Wael Louis

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

Copyright © 2017 by Wael Louis
Abstract

Online and Continuous Electrocardiogram (ECG) Biometric System

Wael Louis
Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto
2017

Online and continuous biometric system revolves around continuously monitoring identity of subjects using biometrics while using current and past observations only. Electrocardiogram (ECG) signal is prone to noise interference and is a slow signal to acquire. First, we developed an online abnormal electrocardiogram heartbeat detection and removal using one-class Gaussian mixture model of two components. This outlier removal method was implemented in a biometric system and was examined on a 1,012 fingertip acquired ECG signals database. The biometric system had an equal error rate (EER) of 5.94% in comparison to 12.30% in a state-of-the-art approach. Due to eliminating noisy heartbeats, the system may suffer from data imbalance problem, and for that we proposed a method to synthesize data. Data synthesis was based on the assumption that ECG heartbeats exhibit a multivariate normal distribution. When small sample size dataset was simulated and examined in a biometric system, EER of 6.71% was achieved in comparison to 9.35% to the same system but without data synthesis.

It was desired to increase biometric system robustness and design a continuous authentication system; hence, a novel feature extraction and a unique continuous authentication strategy were proposed. One-Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP), an online feature extraction for one-dimensional signals was designed, and it was incorporated with sequential sampling to establish a continuous authentication system. This system adaptively updated decision thresholds and sample size during
run-time. 1DMRLBP accounts for observations’ temporal changes and has a mechanism to extract one feature vector that represents multiple observations. 1DMRLBP also accounts for quantization error, tolerates noise, and extracts local and global signal morphology. When 1DMRLBP was applied on the 1,012 fingertip single session subjects database, an EER of 7.89% was achieved in comparison to 12.30% to a state-of-the-art work. Also, an EER of 10.10% was resulted when 1DMRLBP was applied to the 82 multiple sessions database. Experiments showed that using 1DMRLBP improved EER by 15% when compared to a biometric system that was based on raw time-samples. Lastly, when 1DMRLBP was implemented with sequential sampling to achieve a continuous authentication system, 0.39% false rejection rate and 1.57% false acceptance rate were achieved.
Dedication

To my better half, Shahad
Acknowledgements

“With man this is impossible, but with God all things are possible.” Matthew 19:26

First and foremost, I would like to thank Prof. Dimitrios Hatzinakos for being supportive through all the ups and downs in this journey. I am lucky to have the opportunity to work with such an intelligent and a kind person, and I hope to continue working with him many years to come. I also would like to thank the committee members and examiners for their feedback and insight: Prof. Deepa Kundur, Prof. Draper Stark, Prof. Parham Aarabi, Prof. Alberto Leon-Garcia, Prof. Ravi Adve, and my external examiner, Prof. Patrizio Campisi.

Special thanks go to a great unforgettable mentor and a role model who I leaned from and admired immensely, late Prof. Tas Venetsanopoulos. He and his life lessons will be remembered for lifetime. I also would like to thank Prof. Ling Guan for making himself available whenever I needed him. He was the one to introduce me to the realm of academic research in multimedia processing.

Kudos to all the staff and students in the ECE Communication Group, especially Mary Stathopoulos, Sahar Javaher Haghighi, and Majid Komeili. I would also like to acknowledge and thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Edward S. Rogers Sr. Department of Electrical & Computer Engineering at the University of Toronto to fund this project.

Now it is time to thank the core of my life. I would like to thank my parents, Wadoud and Linda; and my brothers, Wassim and Walid, for believing in me and for their continuous support and prayers. Mama and Baba, you did for me more than what any child had hoped to get from their parents. Bros, gym and sheesha nights did not go waste. Parents and bros, I do not have enough space to acknowledge the good you have done to me. I am thankful to my in-laws, Nashwan, Thuraya, Rand, Fahad, and Rana. Keeping the best for last, my soulmate, Shahad. I cannot imagine this day would have come without your support. We celebrated the good times together, and you were
the first one to stand beside me in my difficult times (which were many!). You were there when I needed you, and you were always encouraging and ready to help. On these tight situations, you ensured to keep a smile on your face and could always cheer me up. You made watching TV, eating popcorn, going for long walks ritual habits, and I cannot thank you enough for this. With what seems short time of our life together, the list of good things you have done for my sake can go on and on. Everything was entirely altruistic. You are the best wife one could ever imagine to have. Thank you always, I love you.
# Contents

1 Introduction ................................................. 1
   1.1 Medical Signal Biometrics ................................. 4
   1.2 Electrocardiogram (ECG) Signal ............................ 6
   1.3 Motivations for Using ECG Biometrics ......................... 7
   1.4 Challenges .................................................. 9
   1.5 Contributions .............................................. 10
   1.6 Publications ............................................... 14
   1.7 Summary and Organization .................................. 15

2 Related Work .................................................. 16
   2.1 ECG as a Biometric System ................................. 17
      2.1.1 Fiducial Points Based ECG Analysis ....................... 17
      2.1.2 Non-Fiducial Points Based ECG Analysis ................... 22
   2.2 Irregular ECG Heartbeat Detection and Removal ............... 27
      2.2.1 Targeted Irregularity Detection .......................... 32
      2.2.2 ECG Heartbeats Outlier Removal .......................... 37
   2.3 Data Synthesis and Small Sample Size Problem ................. 44
   2.4 Continuous Authentication .................................. 45
   2.5 Chapter Summary ............................................ 50
### Chapter 3: Experimental Set Up

3.1 Database
- 3.1.1 Physikalisch-Technische Bundesanstalt (PTB) Database
- 3.1.2 University of Toronto Database (UofTDB)

3.2 Experimental Set Up
- 3.2.1 Preprocessing
- 3.2.2 Segmentation
- 3.2.3 Classification
- 3.2.4 Medical Application Experiment Set Up
- 3.2.5 Biometric System Experiment Set Up

3.3 Performance Measure

3.4 Summary

### Chapter 4: Online Outlier Removal

4.1 Introduction

4.2 Methodology
- 4.2.1 Gaussian Mixture Model (GMM)
- 4.2.2 Modeling Normal Heartbeats

4.3 Experimentation

4.4 Summary

### Chapter 5: Data Synthesis

5.1 Introduction

5.2 Methodology
- 5.2.1 ECG Heartbeats Synthesis
- 5.2.2 Parallel Classifiers to Reduce Instability

5.3 Experimentation
- 5.3.1 Synthesized ECG Heartbeat Generation
5.3.2 Comparison to Dimensionality Reduction
5.3.3 Parallel Classifier
5.4 Summary

6 1DMRLBP

6.1 Background of Local Binary Patterns
6.1.1 Review of Local Binary Patterns
6.2 Review of Temporal and Spatial Local Binary Patterns
6.3 One-Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP)
   1DMRLBP Accounts for Quantization Error
   1DMRLBP Captures Multiple Resolutions
   1DMRLBP Considers Temporal Changes
   1DMRLBP Achieves Robustness Towards Noise
6.4 Experimentation
   6.4.1 1DMRLBP as a Feature Extraction
   Experiment on PTB Database for Medical Application
   Experiment on UofTDB for Biometrics Application
6.4.2 1DMRLBP with Different Dimensionality Reductions Techniques
6.4.3 1DMRLBP Features Contribution to Accuracy
6.4.4 Capability and Advantage of Temporal Variation Extraction
6.4.5 Comparison to State-of-the-Art
6.5 Summary

7 Continuous Authentication

7.1 Introduction
7.2 Methodology
   7.2.1 Sequential Sampling
7.3 Experimentation
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.3.1 Comparison to State-of-the-Art CA Systems and Influence of Se-</td>
<td>126</td>
</tr>
<tr>
<td>quential Sampling Parameters</td>
<td></td>
</tr>
<tr>
<td>7.3.2 Continuous Authentication in Action</td>
<td>133</td>
</tr>
<tr>
<td>7.4 Summary</td>
<td>134</td>
</tr>
<tr>
<td>8 Conclusion</td>
<td>139</td>
</tr>
<tr>
<td>8.1 Thesis Summary</td>
<td>139</td>
</tr>
<tr>
<td>8.2 Future Work</td>
<td>143</td>
</tr>
<tr>
<td>8.2.1 Data Synthesis</td>
<td>143</td>
</tr>
<tr>
<td>8.2.2 1DMRLBP Parameters Optimization</td>
<td>144</td>
</tr>
<tr>
<td>8.2.3 1DMRLBP Sparsity</td>
<td>144</td>
</tr>
<tr>
<td>8.2.4 Continuous Authentication Model Adaptation</td>
<td>145</td>
</tr>
<tr>
<td>Bibliography</td>
<td>146</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Summary and highlights of the reviewed ECG biometric systems .......................... 31
2.2 Summary and highlights of the reviewed irregular ECG heartbeat detection and removal .................................................. 43
2.3 Summary and highlights of the reviewed continuous authentication methods .......................... 49

4.1 EER for the biometric systems ........................................................................ 73

5.1 Experiment illustrating data imbalance influence on accuracy .......................... 78
5.2 Experiment for 20 real genuine data. Synthetic genuine data were appended with real genuine data in training the classifier .................................................................................. 84
5.3 Experiment for 30 and 60 real genuine observations ........................................ 86
5.4 Dimensionality reduction techniques best performances based on achieving the lowest EER .......................................................................................................................... 87
5.5 Standard deviation of TAR and TRR for biometric systems with and without parallel classifier .................................................................................................................. 89

6.1 1DMRLBP capability to classify ECG heartbeats ............................................. 105
6.2 EER (%) in different settings .............................................................................. 113
6.3 1DMRLBP features contribution to accuracy ...................................................... 114

7.1 Sequential sampling performance with different parameters ........................... 128
7.2 Performance of Ahmed et al. [1] CA system ...................................................... 129
7.3 Performance (%) of 1DMRLBP with n=1 using continuous authentication

methods in Labati et al. [2] .................................................. 130

7.4 Sequential sampling performance AC/LDA [3] .......................... 131
# List of Figures

1.1 Biometric system operational scheme ........................................ 5
1.2 Positions of the ten electrodes in the 12-lead ECG set up. It is regenerated from [4] .......................................................... 7
1.3 A healthy ECG heartbeat illustrating the main characteristic points ... 8
1.4 Proposed online ECG biometric system with the CA system ............ 11

2.1 Some of the distances used in [5] to illustrate the fiducial points based approach analysis ...................................................... 18

3.1 1-lead UofTDB acquisition configuration [6] .............................. 54
3.2 Continuous ECG signal for 5 second duration ............................. 55
3.3 Illustration of ECG signal preprocessing stage ............................ 56
3.4 Illustration of 67 healthy ECG heartbeats for different subjects ...... 57
3.5 Bagging with 50 decision trees .................................................. 58
3.6 Bagging classifier scheme .......................................................... 63

4.1 Example of a *normal* and an *abnormal* heartbeat ...................... 68
4.2 Heartbeats classification error using GMM ................................. 69
4.3 GMM, M=2 model outlier removal ............................................. 70
4.4 Box plot for 5 subjects showing IQR details ............................. 71
4.5 Offline IQR outlier removal example ........................................ 72
4.6 ROC curves for the biometric systems ..................................... 73
5.1 Synthesized data generation from multivariate Gaussian distribution . . . 79
5.2 Parallel classifier scheme ................................................. 82
5.3 ROC curves for the biometric system with different dimensionality reduction techniques .................................................. 88

6.1 Simple LBP feature. After subtracting the surrounding pixels from the centre pixel and taking their \( \text{sign}(x) \) value, a stream of binary bits are concatenated from an arbitrary but consistent position. ......................... 93
6.2 1DLBP used in [7] ............................................................... 94
6.3 Multi-resolution BP with \( d = 4, p = 5 \) .................................................. 98
6.4 1DMRLBP feature extraction depicts the feature extraction mechanism . 100
6.5 An example of some micro-texture patterns that 1DMRLBP features capture101
6.6 Feature extraction with parameters \( d = (100, 200), \ p = (4, 4), \ \Delta = (100, 200), \ w = (500, 400) \) for an ECG heartbeat. ......................... 107
6.7 \( \text{ROC} \) for parameters \( d = (100, 200), \ p = (4, 4), \ \Delta = (100, 200), w = (500, 400) \). .................................................. 108
6.8 1DMRLBP in a verification biometric system .............................. 110
6.9 Multiple sessions versus single session experiment. EER for single session is 7.89% and for multiple sessions is 10.10% .............................. 111
6.10 Full 1DMRLBP feature vector versus dimensionally reduced 1DMRLBP . 112
6.11 1DMRLBP with temporal changes extraction versus 1DMRLBP without temporal changes extraction among observations ......................... 115
6.12 1DMRLBP versus AC/LDA [3] .................................................. 116

7.1 A hypothetical example illustrating how CA works ......................... 122
7.2 Genuine and imposter distances distributions ............................... 123
7.3 Sequential sampling operation illustrating the criteria and showing an example of accepting a sample after requesting 4 observations ......................... 125
7.4 Sequential sampling criteria with different parameters with $\alpha = 0.01$ and $\beta = 0.01$ ........................................... 136

7.5 Sequential sampling criteria with constant means values ........................................... 137

7.6 A simulated CA performance ........................................... 138
## List of Used Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECG</td>
<td>electrocardiogram</td>
</tr>
<tr>
<td>EMG</td>
<td>electromyogram</td>
</tr>
<tr>
<td>EER</td>
<td>equal error rate</td>
</tr>
<tr>
<td>LBP</td>
<td>Local Binary Patterns</td>
</tr>
<tr>
<td>1DLBP</td>
<td>One-Dimensional Local Binary Patterns</td>
</tr>
<tr>
<td>1DMRLBP</td>
<td>One Dimensional Multi-Resolution Local Binary Patterns</td>
</tr>
<tr>
<td>FRR</td>
<td>false rejection rate</td>
</tr>
<tr>
<td>FAR</td>
<td>false acceptance rate</td>
</tr>
<tr>
<td>TAR</td>
<td>true acceptance rate</td>
</tr>
<tr>
<td>TRR</td>
<td>true rejection rate</td>
</tr>
<tr>
<td>GMM</td>
<td>Gaussian mixture model</td>
</tr>
<tr>
<td>IQR</td>
<td>inter-quartile range</td>
</tr>
<tr>
<td>ROC</td>
<td>receiver operating characteristic</td>
</tr>
<tr>
<td>BP</td>
<td>binary patterns</td>
</tr>
<tr>
<td>DBNN</td>
<td>decision-based neural network</td>
</tr>
<tr>
<td>LDA</td>
<td>linear discriminant analysis</td>
</tr>
<tr>
<td>DWT</td>
<td>discrete wavelet transform</td>
</tr>
<tr>
<td>DCT</td>
<td>discrete cosine transform</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>SVM</td>
<td>support vector machine</td>
</tr>
<tr>
<td>PR</td>
<td>performance rate</td>
</tr>
<tr>
<td>GO</td>
<td>genuine observations</td>
</tr>
<tr>
<td>IO</td>
<td>intruder (imposter) observations</td>
</tr>
<tr>
<td>G</td>
<td>number of genuine or regular heartbeats</td>
</tr>
<tr>
<td>I</td>
<td>number of imposter or irregular heartbeats</td>
</tr>
<tr>
<td>nTP</td>
<td>number of true positive</td>
</tr>
<tr>
<td>nTN</td>
<td>number of true negative</td>
</tr>
<tr>
<td>nFP</td>
<td>number of false positive</td>
</tr>
<tr>
<td>nFN</td>
<td>number of false negative</td>
</tr>
<tr>
<td>CA</td>
<td>continuous authentication</td>
</tr>
<tr>
<td>LLE</td>
<td>local linear embedding</td>
</tr>
<tr>
<td>AC</td>
<td>autocorrelation</td>
</tr>
<tr>
<td>QDA</td>
<td>quadratic discriminant analysis</td>
</tr>
<tr>
<td>FFS</td>
<td>forward feature selection</td>
</tr>
<tr>
<td>SFS</td>
<td>sequential forward selection</td>
</tr>
<tr>
<td>IGR</td>
<td>information-gain ratio</td>
</tr>
<tr>
<td>RS</td>
<td>rough sets</td>
</tr>
<tr>
<td>IMF</td>
<td>intrinsic mode function</td>
</tr>
<tr>
<td>EEMD</td>
<td>ensemble empirical mode decomposition</td>
</tr>
<tr>
<td>EMSSD</td>
<td>root mean square of successive differences</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Allowing access to only authorized personnel has been of interest for thousands of years. Physical key is one of the oldest methods that has been used until this day, and its invention dates back to around 2,000 B.C. in Egypt [8]. Many types of authorization methods have been invented over the years starting from the use of keys, face sketches, and fingerprint sketch [9] to more advanced technology such as access cards and biometrics.

Access control systems are based on either something that is remembered such as PIN and password, something that is possessed such as key and access card, or human characteristic such as fingerprint patterns and face shape [10]. Major drawbacks of credentials that are remembered or possessed are transferability and loss. Even though sharing these entities may not be permitted, for instance sharing access cards in a company, but it nonetheless occurs. As a result, more emphasis has been going towards biometrics [11], the field of study that concerns of people’s identity using their physical or behavioral traits [9].

Biometric systems such as face, voice, and fingerprint recognition systems have gained much popularity. These biometric systems suffer from vulnerability since they are easy to falsify. For example: in face recognition, presenting a photograph of subjects may deceive the system, and similar concept applies to speaker recognition when the subject’s voice
is recorded. Also, there are a few works that reproduce subjects fingerprint using wax which can be used to make synthesized fingerprint, and this might be exploited to gain access to a biometric system. In [12], a wide range of biometric traits along with their advantages and disadvantages are presented. There are some properties that can assess the viability of using particular characteristics as a biometric modality [10], and they are:

- Universality: existence in all healthy human beings;
- Distinctiveness: uniqueness for each person;
- Permanence: the signal does not vary significantly within a short period of time. Fingerprint is an example of a biometric system that does not change over time.
- Conductibility: the possibility and easiness to measure and acquire data. Some biometrics traits are intrusive and not easy to extract such as DNA, while others are as easy as taking a picture such as face recognition.

Despite the advancement in biometric systems, there is no one trait that is suitable to all applications. Properties of each biometrics trait should be studied beforehand. Acceptability is an important factor. It represents how acceptable for general public to use a specific biometrics trait, or how easy it is to ask a person to give his/her biometrics. For example, in access control systems, capturing a picture of a subject is more acceptable than extracting his/her DNA. Accuracy, cooperation of the subject, and hygiene are some of the other factors that play a role in choosing a biometric modality. Different biometrics traits have different accuracy expectation. While fingerprint can be argued to be more accurate than face recognition, yet it needs the cooperation of the subject (active biometrics) to swipe his/her finger on a machine. On the other hand, cooperation may be required (active biometrics) or may not be required (passive biometrics) in face
recognition, for instance access control versus people recognition in a crowd, respectively. Some biometric systems also require touching a sensor which can raise hygiene concerns. After all, all these factors should be weighed to decide on what biometric trait to use.

Most biometric systems consist of four sections [12]:

**Acquisition module:** it is an acquisition device to capture the desired biometrics signal e.g. camera for face recognition, electrodes for electrocardiogram, stethoscope for phonocardiogram, capacitive sensors for some types of fingerprint, and other acquisition sensors.

**Feature extraction module:** it is used to find characteristic and discriminant properties of the acquired signal. In general, feature extraction for \( N \) observations\(^1\), \((X_n)_{n=1}^N\), is transferred to the feature domain \((x_n)_{n=1}^N\), where \( x_n \in \mathbb{R}^k \) is a feature vector of \( k \) dimensions. Features should be *discriminative* in a sense that they attempt to make the signal it is extracted from unique to the subject. Feature vectors should be similar for intra-subjects and different for inter-subjects. They should also be *fast to extract* in order to be applicable to practical applications. Sometimes, feature extraction are not used at all; raw samples of the signals are used instead.

**Classification module:** it is a function \( g(\cdot) : \mathbb{R}^k \rightarrow \mathbb{R} \) that separates, using a risk function, feature vectors into classes. Classifiers are widely examined in biometric systems to find the one that achieves the least classification error. Perfect classification is often infeasible [13]. Difficulties in designing classifiers are related to how well the feature extraction module perform.

**Subjects repository:** it is a database where models for different subjects are stored during a stage called the enrollment stage. Enrollment stage (interchangeably called

---

\(^1\) An electrocardiogram signal observation in this thesis corresponds to a single period of electrocardiogram heartbeat as shown in Figure 1.3.
training stage), is the stage where subjects data are acquired, modeled, then stored in the repository to be used later in the operating mode.

Biometric systems operate in two modes: verification (interchangeably called authentication) or identification. In the verification mode, the system is a binary pattern recognition problem where the output is to accept or reject a claimed identity; hence, it is a 1-to-1 matching. On the other hand, the identification mode is a multi-class pattern recognition problem that tends to find the identity of a subject; it is a 1-to-N matching. Every subject’s model is stored in a repository constructed in a stage called the enrollment stage. Figure 1.1 depicts biometric system modes of operation. However, Figure 1.1 is a more accurate depiction of when the subject provides one observation at a time i.e. in the case of fingerprint. In the case where more than one observation is provided, fusion or decision-making strategy has to be deployed.

There are several methods in the literature that research making a decision from multiple observations in biometric systems [14], but essentially they can be divided into three main streams: feature level fusion, takes place after extracting the feature vector; matching-score level fusion, occurs after each classifier gives a proximity of the resemblance between the feature vector and a specific class; decision level fusion, happens after classifiers made decisions to accept or reject a hypothesis.

1.1 Medical Signal Biometrics

Medical signals have been used for medical purposes for decades such as electrocardiogram (ECG), heart electrical activities; phonocardiogram (PCG), heart sound; electroencephalogram (EEG), brain electrical signal; photoplethysmogram (PPG), organ’s volumetric measure; electromyogram (EMG), muscle signal; and many other biomedical signals that can be extracted from human’s body. In the last decade, some of these signals, namely EMG [15], ECG [16], PCG [11], EEG [17], and PPG [18] were studied as
biometrics.

Medical signals are generated from sophisticated human body functions; hence, they are difficult to replicate and spoof. Also, they can measure liveness of subjects. In other words, they ensure the signal is generated from a human and not from a machine. This is one of the features that sets them apart from other biometrics. Many of the extensively examined biometric traits (i.e. fingerprint, face, iris, and voice) require liveness detection to rule-out signal circumvention. For instance, the biometric system needs to know the face image is not from a picture, the voice is not from a recorder, the iris is not printed on contact lenses, and finally the fingerprint is not from a latex. Some medical signals do not provide adequate subject separation accuracy to be implemented as a standalone system, yet they can still be used in multi-modal system with other biometric traits to increase performance and reduce vulnerability.
1.2 Electrocardiogram (ECG) Signal

In year 1887, the earliest known practical ECG was developed by Augustus Waller, a physiologist [19]. Since then, it has been widely used worldwide to diagnose heart problems.

ECG is a recording of the electrical activity in the heart. In a nutshell, ECG signal is generated by processes called depolarization and re-polarization sequences of the heart. The depolarized signal flow through the heart’s tissues. These sequences are not random, but they propagate in specific patterns over the heart tissues [20]. They are amplified and detected by an ECG device [21].

ECG signal is acquired using electrodes connected to the surface of the skin. The most common configuration of these electrodes is the 12-lead ECG. It uses ten electrodes, six of them are connected to the chest and the other four electrodes are connected to the limbs. Misplacing the electrodes affect the acquired ECG morphology; hence, a protocol is followed to connect these electrodes [21]. Such configuration is called 12-lead ECG because 12 signals are recorded based on different combinations from these ten electrodes. Figure 1.2 illustrates the placements of the ten electrodes on the body. Healthy ECG heartbeat consists of six main characteristic points, P, Q, R, S, T, and U. These points are shown in Figure 1.3.

Each characteristic point is a result of physiological heart depolarization and repolarization. The generation of these points is out of the scope of this thesis, but an in-depth discussion about the source of these points can be found in [21, 20]. Each characteristic point has its own waveform, and a section of an ECG heartbeat is named by which waveform represents it. For instance, QRS waveform corresponds to the duration of an ECG heartbeat that onsets at the beginning of the Q waveform and offsets at the end of the S waveform. In medicine, however, there is a standard criterion to read a QRS complex. It is localized such that the first deflection is downward, and it is called the Q
Figure 1.2: Positions of the ten electrodes in the 12-lead ECG set up. It is regenerated from [4]

wave, following that, if the second deflection is upward, this is called an R wave. If there is an upward trend after the R wave then that is the S wave [21].

1.3 Motivations for Using ECG Biometrics

Tele-medicine is a method to provide medical assistance remotely in areas where medical assistance is not closely available such as third world countries and low populated areas. Assistance includes diagnostic, writing prescription, and all sorts of medical help. Implementing tele-medicine greatly reduces costs of building hospitals and sending medical staff. As a result, a system that can protect and verify the identity of the patients’ data is needed. ECG signal is the most collected signal in tele-medicine [22]; hence, this
collected ECG signal can be used to identify subjects along with utilizing it for other medical tasks.

ECG signal is widely used and studied worldwide to diagnose heart problems. As a result, extensive knowledge has been developed in it. Its usefulness and widespread raised competition among manufacturing companies, and inexpensive sensing devices to acquire the signal have been produced. Due to these factors, ECG signal can be an inexpensive biometric system to deploy.

Research suggests that ECG signal has enough discriminative features to differentiate subjects. Based on the explained biometric properties: universality, distinctiveness, permanence, conductibility, liveliness detection, and invulnerably, they all apply to it. Also, continuous authentication is another advantage that can be achieved since ECG
signal is a quasi-periodic signal. Continuous authentication is a system that continuously monitors the identity of the subject who is proving his/her signal. Hence, continuous authentication finds imposter subjects who wrongfully try to access the system. Not all biometric systems have the continuous authentication capability. For instance, fingerprint cannot be used for continuous authentication because only one data observation can be collected when the person swipe his/her finger.

In summary, ECG signal is hard to spoof, available in all living humans, acquired using inexpensive sensors, and capable of providing person’s clinical status information. These features make ECG signal an appealing signal to investigate as a biometric modality.

1.4 Challenges

Several challenges persist with choosing ECG signal as a biometric system, main ones are as follow:

**ECG irregularities:** Human ECG signal changes based on emotion, diet, or physical exercise [23]. It could also change with diseases and position of the electrodes [21]. A comprehensive list of such benign artifacts are in [20, 24, 25, 26].

**Artifacts in ECG acquisition:** Several signals interferes with ECG signal acquisition, and to name a few:

**Electromyogram interference:** It is an electrical signal due to muscles contractions. It interferes with ECG signals during acquisition.

**Power-line interference:** Depending on the country the ECG is recorded at, 50Hz or 60Hz interference is expected within the acquired ECG signal.

**Baseline wander interference:** Due to respiration, a baseline wander of 0.15-0.30Hz [20] interferes with the ECG signal.
Chapter 1. Introduction

**Contact noise:** Acquisition electrode movement produces noise. It has frequency range between 1-10Hz.

**Relatively slow collection:** ECG signal has a period of an average rate of 1-1.5 heartbeat/second. It is a slow signal if compared to other biometric systems signals such as face in video face recognition that may stream at a rate of 30 frames/second. As a result, for a specific time-frame (e.g. 10 seconds), significantly less number of observations are collected from ECG signal than from a video of faces. The challenge here is that if we, for example, require 100 observations to enroll a subject, and the period for ECG heartbeat is 1.5 seconds, then we need around 2.5 minutes of ECG signal data acquisition, in comparison to 3-5 seconds in video face recognition, and this causes inconvenience in several situations.

**Segment size:** This is more of a general challenge for most biometric systems when multiple observations are provided. Segment size in this thesis is referred to the number of observations the classifier needs to make an acceptable decision. Intuitively, bigger sample size provides more information about the data than a small sample size, yet collecting it is expensive. Hence, there is a trade-off between robustness and accuracy.

### 1.5 Contributions

It is desired to design a robust ECG biometric system that can tolerate ECG artifacts. This biometric system is capable of making a decision using a small number of observations, and it can continuously detect intruders attempts within a short amount of time. This thesis revolves around four phases that can suggest a possible solution to such applications. Figure 1.4 illustrates the overall system design. The phases are:

**Signal quality measure through irregular ECG heartbeat detection:** when ECG
heartbeat data are acquired from a subject, some of the characteristic points are localized. This phase decides whether the heartbeat is regular or irregular heartbeat. In other words, this phase acts as an outlier detection and removal stage. It removes signals that do not obey healthy ECG heartbeat morphology. It is worth mentioning that despite a person’s ECG heartbeat variates [27] with different activities, it stays as a regular heartbeat cycle. This phase corresponds to the QRS detection and the observations processing blocks in Figure 1.4.

This phase implemented an online outlier removal. It is online in a sense that, first, it does not require prior knowledge about the examined subject, and second, it uses previous and current information only to make a decision. We first applied a pre-processing stage to remove some types of noise, then we modeled normal heartbeats using a Gaussian mixture model (GMM), which is used to assess, thus, eliminate abnormal heartbeats in real-time. We utilized a one-class classifier concept. The term normal refers to ECG heartbeats that resemble Figure 1.3, while the term abnormal refers to any signal that is visually significantly different from normal heartbeats (i.e. a signal that does not have characteristic points).

**Synthesize dataset:** Since ECG signal is a slow acquisition signal, we proposed a
method to synthesize ECG heartbeats in order to improve results. This is based on the hypothesis that the more the training data the higher the biometric system accuracy assuming the model does not overfit the training data. This corresponds to the observations processing block in Figure 1.4.

After obtaining a promising results with outlier removal using GMM, the ECG signal Gaussianity was studied. We hypothesized that ECG heartbeats exhibit a multivariate Gaussian distribution. However, the influence of internal and external factors deviate the model from Gaussianity. It was presented that an ECG signal can exhibit a multivariate normal distribution with more than 20 dimensions. Statistical approaches of feature selection and Gaussianity examination were conducted to find these features. Such findings along with the challenge of dealing with small sample size have motivated us to generate synthesized data to improve performance. Each subject data was modeled as a multivariate Gaussian distribution, and synthesized data were drawn randomly from the distribution.

**Design high discriminative features for biometric application:** In this phase we proposed an online feature extraction that can tolerate quantization, scaling, shifting, and misalignment noise; and tolerate abnormal ECG heartbeats. It is online feature extraction and has the capability to capture temporal variations among observations, extracts one feature vector regardless of the number of observations it extracts it from, and preserves morphology of ECG heartbeats. This feature type is called One Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP). This phase corresponds to the feature extraction module in Figure 1.4.

1DMRLBP was inspired by the two dimensional Local Binary Patterns which is successful in image processing, especially in object detection. The image based Local Binary Patterns were modified and enhanced to tolerate all the issues mentioned above. Thorough experiments were conducted on 1DMRLBP to have a solid
understanding of these features and their capability. Experiments suggested that 1DMRLBP has outperformed state-of-the-art features.

**Continuous authentication:** This is the fourth phase. It is a statistical method that was optimized to make a biometric decision with minimum number of observations. This method detects intruders who attempt to access the system. Sequential sampling was proposed to be used for this task. Sequential sampling updates classifiers decision thresholds dynamically based on previous observations classification confidences. This phase corresponds to the continuous authentication block in Figure 1.4.

The proposed continuous authentication system was comprised of 1DMRLBP features and sequential sampling together. Sequential sampling that we proposed to use, was based on two distributions means [28], one distribution was for the genuine subject and the other distribution was for the imposter subject. The distributions were calculated from classifier decision confidences. Based on these statistics, optimized criteria were established to accept or reject an observation. The only parameters that change the criteria set up are the desired Type I and Type II errors. One other property of the proposed continuous authentication system is that it did not require a sophisticated decision-making algorithm because it was based on 1DMRLBP features which extracted one feature vector from multiple observations. Several comparisons and investigations were conducted experimenting the proposed continuous authentication system, and to name a few, we compared this continuous authentication system to another state-of-the-art work that used sequential sampling, a state-of-the-art work that did not use sequential sampling, a continuous authentication with sequential sampling but was based on feature extraction other than 1DMRLBP.

Despite the fact that the main purpose in this thesis is biometric purposes, the contri-
butions and techniques may also be applied to medical diagnosis. For instance, irregular ECG heartbeats detection can be a lifesaver to detect some heart irregularities in early stages which otherwise may get advanced and lead to a lethal consequences.

1.6 Publications

Parts of this thesis has been submitted and published in refereed journals and conferences

**Journal papers:**


**Conference papers:**


1.7 Summary and Organization

ECG signal is used to diagnose heart problems since the beginning of the last century. Recently, ECG signal has been used for biometric purposes where the identity of a subject is to be identified or verified. The slowness of ECG signal period, noise, and variability impose challenges against deploying ECG based biometric systems for practical online applications. Thus, a robust and practical ECG biometric system is the core of this thesis.

In this chapter we introduced reasons explaining the advantages of using biometrics versus traditional authentication methods. We narrowed down the subject of biometrics to the pros of using medical signals biometrics, then we further narrowed the scope of research to reasons behind using ECG signal biometrics in particular. Brief physiology of ECG signal was described. Challenges, contributions of this thesis, and publications of where parts of this thesis have appeared at were reported as well.

This thesis is organized as follows: review of works related to the contributions is introduced in Chapter 2. Chapter 3 presents the methods of evaluation used to measure the effectiveness of the suggested approaches, the databases used, and experimentations set up. Chapter 4 explains the outlier removal contribution. Chapter 5 proposes the data synthesis approach and reports experiments results, and Chapter 6 delves into the main feature type used in this thesis, 1DMRLBP, along with presenting experiments showing the strength of this type of features. Chapter 7 describes continuous authentication contribution and its experimentations. Lastly, Chapter 8 concludes this thesis and suggests possible future work.
Chapter 2

Related Work

One of the earliest works that examined ECG for biometric system was in 1977 [29]. However, the need for biometric systems based on biomedical signal did not catch much attention until the millennium [30]. Many configurations for ECG acquisitions have been deployed but 12-lead ECG configuration is the most popular technique. Majority of the papers in the literature examined one reading out of the 12 readings [16] in the 12-lead ECG configuration. Simplicity of not dealing with 12 signals fusion is the main reason for selecting a single signal; however, some works examined all 12 readings as in [31, 32, 33]. In this chapter, literature review about ECG signals in areas pertinent to this thesis is presented. In Section 2.1, research approaches that used ECG signal as biometrics are explained. In Section 2.2, irregular ECG heartbeats detection is reviewed. Section 2.2 along with Section 2.2.2 review works that detect only or detect then remove irregular ECG heartbeats. Lastly, continuous authentication review is presented in Section 2.4. Towards the end of each subsequent section, we summarize the reviewed techniques and state, when applicable, the reasons hindered using such techniques for our desired application.
2.1 ECG as a Biometric System

This section demonstrates literature approaches toward using ECG signal biometrics for subject recognition. Healthy ECG heartbeat has six characteristic or fiducial points, namely P, Q, R, S, T, and U. This is illustrated in Figure 1.3. Based on these fiducial points, two mainstreams of ECG analyses have been investigated in the literature: fiducial and non-fiducial points based analyses.

ECG heartbeats are segmented and aligned to have a persistent feature extraction. Most of the techniques, whether they are fiducial or non-fiducial points based approaches do signal segmentation and alignment. Some exceptions exist such as the works in [3, 32, 34, 35].

2.1.1 Fiducial Points Based ECG Analysis

In fiducial points based approaches, features are usually extracted from the six characteristic points in Figure 1.3. Features can be distances between points, angles, areas, amplitudes, etc. Extracting features from fiducial points is the most intuitive approach since this is how practitioners assess ECG heartbeats in the medical area [36]. Figure 2.1 shows some of the 21 distances used in [5].

One of the pioneer works in ECG biometrics is in [30]. The authors presented the possibility of classifying 20 subjects with high accuracy. Signals from 12 leads were used and segmented. From each ECG heartbeat, 30 features of temporal and amplitude attributes were extracted from the fiducial points (P, Q, R, S, T, and U) along with their onsets and offsets. For the 12 leads, 360 features were extracted \((30 \times 12)\). It was suggested that one lead was adequate to find the discriminative features of a subject. This suggestion was based on the fact that there was a strong correlation among different leads for a specific feature. The authors also experimented using six leads only from the limbs and ignoring the chest leads. This is because limb leads are easier to attach to the body.
Chapter 2. Related Work

Figure 2.1: Some of the distances used in [5] to illustrate the fiducial points based approach analysis.

Soft independent modeling of class analogy (SIMCA) was used for classification. The best result was obtained when only 10 features out of the 360 features were used.

The work in [22] is also an early successful work in the area of fiducial points based ECG signal recognition. Temporal information of only 4 durations, namely PQ, P, QRS, and QT were considered. The authors suggested that these durations are not affected by RR interval variability and electrodes conditions. Second derivative was used to localize the characteristic points. All combinations of two from the four attributes were considered in the experiments to find the best discriminative features. Average, variance, and covariance values were used to generate Mahalanobis generalization distance coefficients. The smallest Mahalanobis distance corresponded to the identity of the subject. In their experiments, it was noticed that QRS and QT intervals contributed to the highest discriminative features while P interval had the lowest discriminative power. The authors suggested using amplitude and ratio of amplitudes since it might also be invariant to ECG irregularities and electrode conditions.
In [37], preprocessing stage was applied to remove baseline wander, DC offset, power-line interference, and high frequency interference. Features were extracted from the Q, R, S, and T fiducial points. To reduce heart variabilities, heartbeats were normalized by the QT duration. Two stages of classification were examined: template matching and decision-based neural network (DBNN). Template matching was based on correlation measure. If the correlation coefficient $≤ 0.85$, then it was decided as a mismatch otherwise the feature vector was passed to the DBNN. The authors suggested that this approach is computationally simple, has reasonable training time, is capable of offline training, has a pre-screening stage using template matching which increased accuracy, and has the capability of achieving perfect identification rate on the examined 20 subjects.

In [38], the peak points of P, Q, R, S, and T were detected. The detection method started by locating the R peak, then Q and S peaks were localized as the the local minima around the R peak. The P and T peaks were detected by locating maxima within a range of time-samples around the R peak. The authors acknowledged the presence of baseline wander and used high-pass IIR filter with cutoff frequency of 5Hz to eliminate it. Each heartbeat amplitude was normalized by diving every sample by the R peak value. RR interval was normalized in two methods. First approach was consisted of normalizing RR duration to 100 samples by interpolating every sample between the RR waveform. The second method of normalization was implemented by fixing the places for the P, Q, S, and T peaks, then the method suggested interpolating the samples among them. Two signals were considered in the feature extraction, the original one and the one with normalized RR interval. On both signals, wavelet transform was applied. In particular, Mexican hat wavelet was used. Quadratic discriminant analysis (QDA) was used for classification.

In [39], band-pass filter of low order was used to preserve information from 1.1-40Hz while eliminating the noise at 0.06Hz and 60Hz (i.e. baseline wander and power-line interferences, respectively). The fiducial points P, Q, R, S, and T were extracted along with other four points T’ (end of T wave), S’ (end of S wave), P’ (end of P wave),
and L’ (start of P wave). Hence, the authors used a total of nine points. Out of these nine characteristic points, fifteen durations were extracted to create the feature vector. Twelve out of the fifteen attributes were used in most of the experiments. Only temporal information was gathered since it was claimed that temporal information did not change if the sensor was misplaced or moved, unlike amplitude. The R peak was used to align ECG heartbeats. This work examined the effect of the position of the sensor and the emotional state of the subject. It was suggested that normalization by the distance L’T’ made features invariant to anxiety state. Linear discriminate analysis (LDA) was considered as a classification stage then majority voting was applied to assign individuals to heartbeats data.

In [40], temporal and spectral features were investigated. Discrete wavelet transform (DWT) was applied to find fiducial points, and the offset and onset of an ECG heartbeats. DWT was also used to eliminate noise. The quadratic spline wavelet for five scales were used. Normalization of amplitude and temporal features were employed to reduce the effect of heart rate irregularities. Thirty-four features from time intervals and amplitudes were extracted. To reduce dimensionality, thirteen out of the thirty four features were extracted using Principal component analysis (PCA). The authors claimed that they improved performance and reduced complexity using dimensionality reduction. Also, multi-class support vector machine (SVM) was used in the classification stage.

In [23], multiple filtration stages were used as a first stage. Heartbeat averaging, wavelet, least squares modeling, and Fourier band-pass filtering were the filtration stages. The purpose was to design a stable filter that is applicable to all subjects and capable of retaining intrinsic signal while rejecting noise. Band-pass filter with cutoff frequencies of 0.2-40Hz was utilized. A low order filter was used to reduce the edge effect. To locate the R peak, maximum variance of 0.2 second intervals was examined. The other peaks, P, T, Q, and S, were localized in a relative position to the R peak. Similar to [39], T’, S’, P’, and L’ fiducial points were localized. These peaks along with RR interval
were the nine features that constructed the feature vector. These features were modeled as a multivariate Gaussian distribution, and decisions were made using sequential sampling criterion examining the claimed identity (null hypothesis) and the closest imposter identity (alternative hypothesis).

In [5], Butterworth band-pass filter with cutoff frequencies of 1-40Hz was used to reduce the effect of noise. QRS complex was detected and aligned by the R peak. The detected ECG segment was considered as a window of 800 milliseconds centered at the R peak. Temporal attributes, similar to [23, 39], were utilized, but the features were normalized by P’T’ duration to reduce variability. Along with the temporal attributes, six amplitude attributes were extracted. LDA was experimented for classification. The paper suggested that the six amplitude attributes significantly improved the results. Apart from these attributes, PCA and LDA were exploited to extract holistic patterns in the signal.

Similar to [5], the work in [41] incorporated amplitude with temporal features to construct 24 features. In [41], it was suggested that temporal features do not extract good intra-subject features. Similar conclusion was obtained in [5] where it was suggested that amplitude features are more discriminative than temporal features. Sequential forward selection (SFS) algorithm was employed to find a subset of features that achieved a predefined training error. The authors selected nine features out of the 24 features. They observed that nine feature achieved the best accuracy. Increasing the number did not improve the results. Also, Mahalanobis distance was applied as a classifier.

The work in [42] followed the same regular stages of fiducial point based approaches. It started with ECG filtration then detected and aligned fiducial points. These steps were followed by feature extraction. Three angle attributes were incorporated along with temporal and amplitude features. A total of 20 attributes were considered. Correlation as a matching method was used to decide on the identity of subjects. If two heartbeats had a correlation of more than a specified threshold then it was accepted, otherwise, it
was rejected.

In [43], several feature reduction methods were examined, namely PCA, LDA, information-gain ratio (IGR), and rough sets (RS). For preprocessing, Butterworth band-pass filter with cutoff frequencies of 0.5-40Hz was applied. Fiducial points along with their onsets and offsets were detected. Details on information on the fiducial point detection used in [43] is presented in [44]. Twenty-eight attributes were extracted from the fiducial points, where 19 of them were temporal, six of them were amplitude, and three of them were angle attributes. Normalization by the length of heartbeat duration was applied to resist variability effect. Furthermore, amplitude attributes were normalized by the R peak amplitude to avoid signal attenuation. Radial basis functions neural network was exploited as a classifier due to it’s rapid training and simplicity. The authors concluded that amplitude and angle features were more discriminative than temporal features. IGR and RS performed better than PCA and LDA. RS performed the best of all methods.

The main disadvantage using fiducial point analysis is the need to locate fiducial points in an ECG heartbeats. Error in accurately locating these fiducial points sabotages the biometric system performance.

2.1.2 Non-Fiducial Points Based ECG Analysis

Non-fiducial points based analysis technique is a holistic approach that considers the ECG signal or the isolated ECG heartbeat as a whole. Non-fiducial points based approach, it does not mean fiducial points are not extracted. Majority of the work in non-fiducial based approaches extract some fiducial points for the sake of aligning and segmenting ECG heartbeats. Hence, if the fiducial points were used to find the features i.e. durations, angles, amplitudes then that is a fiducial point based approach; otherwise, it is a non-fiducial point based approach [45].

Non-fiducial point based analysis can be further subdivided into two mainstreams: one that needs to segment and/or align the data such as the works in [46, 47, 48] while the
other does not require any information about the signal such as the works in [3, 35, 49].

The approach in [50] applied Lyapunov exponents spectrum and correlation dimension to capture the indexes of chaotic ECG signal. Before this research, chaotic theory was used to distinguish between healthy and non-healthy signals. In this paper, signal from one lead was used. From the chaotic ECG extractor features, a threshold was set to decide whether the signal was an outlier or not. If it was not an outlier, back propagation neural network was used for classification.

In [51], DWT was used to de-noise the ECG signal. The authors acknowledged that ECG signal is a non-stationary signal so time-frequency representation outperforms other filtration techniques such as digital filters, adaptive de-noising methods, and frequency domain analyses. In the DWT de-noising, a mother wavelet and its levels were selected statistically such that it achieved the lowest reconstruction error between ECG signal and the coefficients of the wavelet in a level. Daubechies, symlet and coiflet wavelets were tried. The best result was achieved using coiflet with 4 levels. Autocorrelation coefficients were considered as the features. Afterwards, linear and non-linear dimensionality reduction techniques were experimented. In particular, PCA, kernel principal component analysis (KPCA), and LDA were used. The dimensionally reduced feature vector was classified using multi-class SVM with one-against-all approach.

The work in [46] applied DWT directly to raw ECG signals. Quadratic spline mother wavelet was used. No other preprocessing stage was suggested since filtration was implicitly performed while using DWT. QRS, T, and P waves detection method was motivated by the work in [52]. They were detected by investigating different wavelet scales. With that, the onset and offset of an ECG heartbeat was localized. For normalization, the section of a signal that started from the end of the QRS waveform to the end of the T waveform was normalized to 120 milliseconds since it was suggested that only this duration is heart-rate dependable. The entire heartbeat was then re-sampled to have 850 millisecond duration. It was concluded that such normalization made the heartbeat
invariant to emotion irregularities. In this paper, correlation was used for classification.

The work in [47] is one of the early works that collected ECG signals from fingers. A notch filter was used to reduce the power-line noise. Fiducial points P, Q, R, S, and T were detected using the algorithm proposed in [53]. The alignment was conducted using cross-correlation. Any segmented heartbeat that had a cross-correlation coefficient below one standard deviation of the mean correlation coefficient was discarded. All remaining ECG heartbeats were fused by generating a single averaged ECG heartbeat. This procedure reduced low frequency drifts and additive noise. For classification, wavelet distance measure was exploited. Along with that, residual difference and correlation coefficient were examined. The authors argued and speculated that ECG biometrics performance degraded on larger datasets (they used a database of 50 subjects); hence, they proposed using ECG biometrics as a supplement to another biometric system such as fingerprint biometrics.

DWT was also examined in [48]. Preprocessing was based on two types of filters: high-pass and low-pass filters. High-pass filter with cutoff frequency of 0.5Hz was used to eliminate baseline drift while a low-pass filter with cutoff frequency of 45Hz was applied to eliminate high frequency noise such as power-line interference. QRS waveform detection was achieved using a method described in [54]. The ECG heartbeat segment was considered of a length of 128 samples from data sampled at 250Hz. Each segment corresponded to 43 time-samples before the R peak and 84 time-samples after the R peak. Each four consecutive segments were utilized for feature extraction. Wavelet coefficients were extracted from each four segments (512 time-samples). Haar wavelet with nine levels decomposition was explored to calculate wavelet coefficients. For classification, Euclidean distance was used.

While all previous works required segmentation and alignment, the work in [35] is one of the earliest works that did not require any information about ECG fiducial points. The authors acknowledged that there is no definitive and universally accepted rule about
the onset and offset of an ECG heartbeat. In [35], a band-pass filter with cutoff frequencies of 0.5-40Hz was first applied to the ECG signal. Afterwards, autocorrelation of a window length of more than the duration of one heartbeat was calculated. The autocorrelation was normalized by the maximum correlation value to cancel biasing factors. Discrete cosine transform (DCT) was applied on the autocorrelation coefficients for dimensionality reduction. Normalized Gaussian log likelihood and Euclidean distance were used for classification. A perfect identification was achieved with the normalized Gaussian log likelihood distance. In [5], even if the authors have proposed fiducial points based approach as explained in the previous subsection, this same work has proposed a non-fiducial points based approach. Similar work to [35] was proposed. Autocorrelation with window of a length longer than the heartbeat period was applied. DCT was used on the lags of the autocorrelation, and its coefficients were used for classification.

The work in [55] is also oriented in the same direction as the work in [35]. It applied a Butterworth filter with a cutoff frequencies of 1-40Hz. The ECG signal was segmented into non-overlapping windows. For these windows, autocorrelation was calculated and was normalized by the maximum value. DCT and LDA were applied for dimensionality reduction. LDA outperformed DCT since it was supervised (i.e. the subjects classes are provided while training). Euclidean distance was used for classification.

Fusion of temporal and cepstral information was introduced in [49]. Normalization was performed on the temporal attributes for preprocessing. Also band-pass filter with cutoff frequencies of 0.05-40Hz was used. Hermite polynomial examples were exploited in the time-domain as features and SVM was used for classification. Moreover, cepstral mean subtraction and cepstral variance normalization were also used for feature extraction. Heteroscedastic discriminant analysis [56] was applied for dimensionality reduction then Gaussian mixture model was used for classification. Fusion at the score level of both types of features suggested promising results.

Ensemble Empirical Mode Decomposition (EEMD) based method was proposed in [57].
For preprocessing, multiple types of noise elimination were applied such as de-trending, wavelet minimax thresholding, and wavelet shrinkage de-noising. For de-trending, time varying band-pass filter was used to eliminate baseline drift and low frequencies. Biorthogonal spline wavelet with 3 levels was used to decompose the ECG signal. Using these wavelets along with minimax thresholding method, high frequency was removed. Also, wavelet shrinkage de-noising was used to reduce the noise resulted from the non-stationarity of ECG signals. In regards to heart rate variability effect, normalization was applied to reduce it. Heartbeats were normalized to have 75 beats/minute. The normalization was based on detecting fiducial points, and those were extracted using biorthogonal spline wavelet. Quality measure was conducted using periodicity transform. After all these stages, the feature vector was constructed as follows: intrinsic mode function (IMF) of the ECG signal, which was estimated using EEMD, was applied to decompose the signal; and, Welch method was used to find the spectral power of the IMF signals. Both IMF components and their spectral power were considered as features. PCA was applied for dimensionality reduction and to reduce computational complexity. KNN using Euclidean distance was examined for classification.

For completeness, some works were oriented towards hybrid approaches where they used both fiducial and non-fiducial points based features. In [45], preprocessing for noise removal was applied using an FIR filter. Also, amplitude normalization was applied. Afterwards, ECG signal was segmented after finding its fiducial points. From the segmented heartbeats, the work in [58] was used for feature extraction. Fifteen features were extracted from every heartbeat using seven-level wavelet transform. The segmented heartbeat were considered as 400 milliseconds around the detected R peak, and two more features were extracted from the RR interval. These two features were the duration from current R peak to previous R peak, and the duration from current R peak to next R peak. At last, SVM was used for classification. Another hybrid work is in [5]. Amplitude and temporal information was extracted from fiducial points, PCA was applied on the raw
time-samples heartbeats to capture a set of non-fiducial features. The feature vector was the concatenation of both feature sets. Wilks’ Lambda for feature selection was deployed then LDA was used for classification.

Lastly, in [59], a bandpass filter with cutoff frequencies of 0.5-45Hz was used to remove baseline wander and power-line interference. The signal was then smoothed with a three-sample width and an eight-point Gaussian window. Segmented heartbeats were extracted using convolution with a window [60]; in other words, using template matching. Fifty-five fiducial points were extracted using the method in [60]. The fiducial points included temporal and other features such as slopes, onset and offset points, durations, maxima and minima, etc. On the other hand, non-fiducial features were extracted by applying DCT on every extracted ECG heartbeat. The pool of features of both fiducial and non-fiducial features had 115 features. Feature forward selection (FFS) was used to select a subset of these features. Several classification methods were experimented on the extracted features. In particular, Gaussian data description, Parzen window data description, SVM, K-means data description and K-centre data description.

The main disadvantage of this approach is data redundancy. In this thesis, we will pursue non-fiducial points based approach by proposing a special type of features, we called One-Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP). These features outperform state-of-the-art work and can tolerate a range of noise sources. Table 2.1 presents a summary of the reviewed works for fiducial and non-fiducial points based approaches.

2.2 Irregular ECG Heartbeat Detection and Removal

Work has been conducted to detect ECG irregularities. Most of the research was intended for medical diagnoses; however, in this thesis, ECG irregularities detection is used as heartbeat quality measure to filter undesired signals for biometric systems. This
section reviews approaches that assisted us to reach our proposed technique in Chapter 4. Approaches can be split into two mainstreams: one that targets a specific type of irregularity then try to process the signal, and the other approach models the regular signal (one-class classifier) and considers everything that is not classified regular heartbeat as an irregular heartbeat then process the signal.
<table>
<thead>
<tr>
<th>Citation</th>
<th>Feature extraction</th>
<th>Classification</th>
<th># sub.</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biel et al. [30]</td>
<td>temporal and amplitude</td>
<td>SIMCA</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td>Kyoso et al. [22]</td>
<td>temporal</td>
<td>Mahalanobis generalization</td>
<td>9</td>
<td>76.1%</td>
</tr>
<tr>
<td>Shen et al. [37]</td>
<td>temporal and amplitude</td>
<td>template matching and</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td>Fernando et al. [38]</td>
<td>wavelet coefficient of aligned and unaligned signals</td>
<td>DBNN</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>Israel et al. [39]</td>
<td>temporal</td>
<td>QDA</td>
<td>29</td>
<td>100%</td>
</tr>
<tr>
<td>Kaveh et al. [40]</td>
<td>temporal and spectral</td>
<td>multi-class SVM</td>
<td>18</td>
<td>98.7%</td>
</tr>
<tr>
<td>John et al. [23]</td>
<td>temporal</td>
<td>sequential sampling</td>
<td>29</td>
<td>100%</td>
</tr>
<tr>
<td>Wang et al. [5]</td>
<td>temporal</td>
<td>LDA</td>
<td>75</td>
<td>*</td>
</tr>
<tr>
<td>Gahi et al. [41]</td>
<td>temporal and amplitude</td>
<td>PCA</td>
<td>29</td>
<td>100%</td>
</tr>
<tr>
<td>Singh et al. [42]</td>
<td>angle, temporal, and amplitude</td>
<td>SFS and Mahalanobis distance</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>correlation coefficient</td>
<td>50</td>
<td>99%</td>
</tr>
<tr>
<td>Authors</td>
<td>Methods</td>
<td>Dimensionality Reduction and Classification</td>
<td>Accuracy</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>Tantawi et al. [43]</td>
<td>angle, temporal, and amplitude</td>
<td>dimensionality reduction and neural network</td>
<td>96±2%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>93.2±4%</td>
<td></td>
</tr>
<tr>
<td>Chen et al. [50]</td>
<td>Lyapunov exponents and correlation dimensions</td>
<td>neural network</td>
<td>≥ 90%</td>
<td></td>
</tr>
<tr>
<td>Hejazi et al. [51]</td>
<td>wavelet and autocorrelation coefficients</td>
<td>dimensionality reduction and SVM</td>
<td>88.18±0.82%**</td>
<td></td>
</tr>
<tr>
<td>Fatemian et al. [46]</td>
<td>wavelet coefficients</td>
<td>correlation coefficients</td>
<td>99.61%</td>
<td></td>
</tr>
<tr>
<td>Chan et al. [47]</td>
<td>average of heartbeats</td>
<td>wavelet distance</td>
<td>89%</td>
<td></td>
</tr>
<tr>
<td>Chiu et al. [48]</td>
<td>wavelet coefficients</td>
<td>Euclidean distance</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Plataniotis et al. [35]</td>
<td>autocorrelation coefficients</td>
<td>DCT and normalized Gaussian log likelihood distance</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Agrafioti et al. [55]</td>
<td>autocorrelation coefficients</td>
<td>LDA and Euclidean distance</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Li et al. [49]</td>
<td>Hermite polynomial in time domain and cepstral features</td>
<td>fusion of SVM and GMM</td>
<td>98.26%</td>
<td></td>
</tr>
<tr>
<td>Zhao et al. [57]</td>
<td>IMF and their spectral power</td>
<td>KNN</td>
<td>95%</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.1: Summary and highlights of ECG biometric systems. *not significantly different from the 29 subjects above it. **this is window recognition rate; subject recognition rate was 40.38±5.08%
2.2.1 Targeted Irregularity Detection

In [61], the authors detected specific heartbeat irregularity. For the task, QRS waveform was first detected using Pan-Tompkins [62] method, then the signal was preprocessed by normalizing waveform and applying filters. A high-pass filter was first applied to remove the DC component, then a band-pass filter with cutoff frequencies of 5-15Hz was used to extract intrinsic ECG signal. A derivative filter was exploited to detect the QRS slope. Lastly, R waveform was detected by applying a moving window integrator. To find an irregularity, atrial fibrillation in specific, the time-domain root mean square of successive differences (RMSSD) [63] was used. Once irregularity was detected, correlation was used to find its position and duration.

The authors in [64] chose to find regular and irregular waveforms using correlation only. The used database had sampling rate of 350Hz. Each heartbeat was segmented to 150 samples. After detecting the R peak, a duration of 49 samples left of the R peak to 100 samples to right of the R peak were considered the heartbeat segment. The algorithm stored reference signals for different irregularities. Cross-correlation was applied between ECG signal and the reference signals to find the highest correlated signal. The $R^2$ measure of the correlation was considered as the distance to decide whether a matching had occurred or not. In other words, this is a template matching technique.

In [31], discriminant analysis was applied to multivariate time series. The wavelet correlations and wavelet variances were used to distinguish between healthy and non-healthy (in other words regular and irregular, respectively) ECG heartbeats. Daubechies mother wavelet with lengths 2, 4, 6, and 8, symmletts mother wavelet of length 8, and coiflets mother wavelet of length 6 were examined. It was shown that using wavelet variances and wavelet correlations together performed significantly better than using wavelet variances only. On the other hand, in some cases, using wavelet correlations alone performed similar to the case when wavelet correlation and variances were used together. In general, classification using linear discriminative analysis outperformed the
QDA. The authors claimed that this work was insensitive to preprocessing, signal length, and sampling frequency, when the examined window is more than 2 second length.

Auto-regression (AR) Burg algorithm parameters were used as features in [65]. Burg algorithm estimates real value AR coefficients using AR coefficients of previous order. AR models are intensively studied in biological signals; hence, it was examined in this paper. C4.5 decision tree algorithm based on ID3 algorithm was examined to classify the data as being healthy or non-healthy (regular or irregular) heartbeats. In designing the decision tree, number of leaves were eight leaves and the tree was of size 15.

DWT is widely used in ECG signal since it achieves a good frequency resolution at low frequencies and a good time resolution at high frequencies [66]. In [66], multiscales of DWT were considered as features. Haar wavelet in particular was used as a mother wavelet since it is the shortest and simplest wavelet among the examined ones. Also, Haar wavelet is capable of locating time-domain signal characteristics. A two-level wavelet decomposition was applied. Morphology, coherence, and energy were the features extracted from the decomposed waveforms. Afterwards, normalization of the feature vector was performed using hyperbolic tangent sigmoid function on the zero-mean and unit standard deviation of each feature. Probabilistic neural network was suggested as a classification scheme for these features. This neural network is a type of the radial basis function, and it uses Gaussian kernel. The ECG heartbeat was classified into one of six classes of irregularities. The authors suggested that the parameters that affected the accuracy of the systems were the smoothing factor and the training sample size. The smoothing factor corresponded to the standard deviation of the Gaussian kernel. The works in [67, 68] also examined wavelet for irregularity classification.

The aforementioned techniques used only one lead; however, there are some works that used more than one lead. In [69], leads I,II, and III were used. This paper captured heartbeat irregularity due to medical issues. The signal was filtered with simple median filter to reduce overshoots from motion artifacts, then average filter was used to reduce
jagged edges. Two stages of polynomial fitting approaches were utilized to eliminate baseline wander: the fourth order polynomial was used to reduce the effect of baseline wander then an eighth order polynomial was applied. Each heartbeat segment consisted of 0.4 seconds. From each segment, a total number of 24 features were extracted from the fiducial points from the three leads, and each of them were normalized. Another stage was dimensionality reduction, and for that, PCA was used to reduce redundancy to 15 dimensions. Finally, Elman neural network was utilized for classification. One of the observations of this method is that it was noticed that most of the redundant features were from lead II.

In [70], the work proposed a simple framework to detect an irregularity. The type of irregularity can be identified if the P waveform is detected precisely, yet identifying the P waveform is not as simple as detecting the QRS waveform, especially in the presence of heartbeat irregularities. Hence, RR intervals were utilized for heartbeats detection. In this paper, the authors resorted to DWT with six levels to extract 24 features. These features trained SVM classifiers to detect heartbeats segments. Furthermore, they extracted another five features which trained other SVM classifiers to detect heartbeat rhythm. From both classifiers (heartbeats and rhythm), another set of classifiers were used to make the decision. The final classifier stage took its input from the rhythm and the heartbeats classifiers, and it consisted of two SVM classifiers. This work was oriented towards performing in real-time in low power devices. The reported classification time was 10.2 milliseconds per heartbeat.

A qualitative approach that uses SVM with Kernel-Adatron learning algorithm was suggested in [71] to find heart irregularity. Experimentation was conducted based on training and testing of one signal then again on two signals. This work mainly compared multi-layered perceptron with back propagation to SVM and found out that SVM was significantly faster and more accurate. The ECG signal was initially preprocessed for baseline wander reduction. Heartbeats were isolated based on RR intervals. This paper
proposed to use three evaluation criteria to make a decision, and those criteria were: training and testing performance plus training time. The authors suggested that using these criteria together eased comparison among other works in the literature.

Another work to find irregular ECG heartbeats and targeted low-processing power devices is in [72]. The authors acknowledged the problem with Holter monitoring devices from the aspect of the tedious work required to process thousands of the collected ECG heartbeats offline. In the proposed system, bandpass filter and QRS waveform detection were first implemented. QRS waveform was detected using Pan-Tompkins in [62]. The bandpass filter was a two stage filtering process consisted of low-pass and high-pass filters. The low-pass filter had a cutoff frequency of 11Hz and a gain of 36 mV/mV. The high-pass filter had a cutoff frequency of 5Hz with a unity gain. The onset and offset of the ECG heartbeats were localized using the method in [73]. Adding band-pass filter introduced Gibbs ringing phenomenon into the Q and S waveforms which led into having an amplitude higher than P and/or T peaks in some occasions which led to P and T detection failure. To reduce Gibbs rings in the filtered signal, an averaging filter with a length of six time-samples was added. Another purpose of the averaging filter was to detect the polarity of P and T peaks to locate their peaks and locations. After detecting fiducial points, durations, slopes, and energy, they were used as features. These features were used to train a feed-forward multilayer perceptron artificial neural network.

The work in [74] is another project that targeted low-processing power DSP. In this work, the authors claimed that the EMG signal is in the range from 30-500Hz while the ECG signal is in the range from 0.5-100Hz. A fourth order Butterworth high-pass filter with cutoff frequency of 30Hz was used. The authors have proposed suppressing the QRS waveform before applying the high-pass filter because the QRS waveform has high frequencies, and they desired low order filter. After applying the high-pass filter, EMG signal was also extracted, and it was subtracted from the filtered ECG signal. The type of irregularity tackled in this paper can be detected from the ST waveform. Hence,
ST waveform was detected and was divided into five equal sections. Slope of the ST waveform was compared to databases to find if irregularities were present. This system was implemented in a DSP device, and apart from the preprocessing steps explained, the baseline wander and the offset were detected and removed by the chopper stabilization and dynamic offset cancellation in the hardware. Recently, several works targeted smartphones as in [75, 76, 77, 78].

In [79], QRS waveform was extracted using difference operation method (DOM). The first step in DOM was noise elimination via preprocessing stage. The power interference was removed by a notch filter, baseline drift was reduced by applying the technique in [80] which is a high-pass filter with 0.5Hz cutoff frequency. EMG signal was removed by a morphological filter [81], which is a unit square wave. Lastly, the motion artifact was reduced with an adaptive filter [82]. The second step intended to find the derivative of the ECG signal then applied a 100Hz low-pass filter to eliminate high frequency. Lastly, using a thresholding and matching mechanisms, the QRS waveform was detected. This technique was fiducial point based technique that extracted nine features based on location, duration, amplitudes, and shapes of the ECG heartbeats. LDA was considered for classification.

Bispectrum analysis was proposed in [83]. It focused on two dimensional third order cumulant Fourier transform. In the bispectrum analysis, we obtain bicoherence function, which is a power spectrum normalized function. It is bispectrum; hence, it is a 2D plane. This plane was divided into eight symmetric areas, and the features were extracted from them. The features consisted of means, variances, and entropies. QRS waveforms were used as the main portion to distinguish normal heartbeats from the types of irregularities the paper desired to find, and features were extracted for each QRS waveform. After extracting the features, three types of classifications were examined: GMM, KNN, and vector distance classifier. Nonparametric KNN and vector distance classifiers outperformed the parametric GMM classifiers, even though the extracted features exhibited
Gaussianity according to Royston’s test, however, they have high variance. Also, despite simplicity of vector distance classifier, which was essentially KNN with K=1, the results were close to each other. With this approach six types of medical irregularities were targeted. Since classification such as simple vector distance achieved a promising results, the authors speculated such approach would be viable for low-level hardware implementation and portable electronics.

Higher order spectra was also used in [84] to distinguish between normal heartbeats and irregular ones. First, QRS waveforms were detected using Pan-Tompkins algorithm [62]. Afterwards, heartbeats were segmented as 150 time-samples around the R peak. Using the higher order spectra, features were extracted from the heartbeats. Independent component analysis was used for data reduction. KNN with K=2, classification and regression tree, and neural network were used to evaluate the performance.

There is a number of papers that tackles ECG irregularities. The works in [85, 86] used multi-layer perceptron, [87] applied Hilbert Transform, [88] applied artificial bee colony algorithm, [89] suggested using prediction by partial matching, [90] used hidden Markov chain, [91] applied sequential Markov chain, and [92] applied Gaussian process classifiers trained with wavelet transform and S-transform.

\section*{2.2.2 ECG Heartbeats Outlier Removal}

Some ECG signal processing is oriented towards finding the quality of an ECG heartbeat, and if a heartbeat is not up-to-par quality then it is considered as an outlier and is removed. The work in [93] measured the quality of ECG signal through three stages. The first stage was a preprocessing stage to remove baseline wanders and high frequency disturbances. Gaussian smoothing filter was utilized to eliminate high frequencies. Gaussian smoothing filter is similar to an FIR low-pass filter but with no hard cutoff frequency. To remove baseline wander, adaptive smoothing filter with a window length equal to the average value of the duration between consecutive R peaks was used. The second stage
consisted of measuring the energy of the signal. After finding the energy of 1,000 manually annotated ECG signals, maximum and minimum energies were calculated. Any signal that had energy greater than the maximum or less than the minimum was considered unacceptable. Also a concavity-based quality measure was used. The QRS complex has a sharp concavity while the rest of the signal does not. By normalizing the concavity, the authors established a measuring criteria. The third stage measured the quality of the signal by applying correlation.

In [3, 6], the features were autocorrelation of a specific number of lags for overlapping windows. If the Euclidean distance between the autocorrelation and the median autocorrelation of all windows for a subject exceeded a threshold, then it was considered as an outlier, and it was eliminated.

The research in [94] performed quality assessment on ECG signal collected from twelve leads. The signals were each of 10 seconds and had several usual ECG noise such as contact noise, motion, and electrode misplacement. The ECG signal was sampled at 500Hz, so the signal had 5000 samples. The authors claimed that the noise exhibited some effect on the ECG signal, and they designed a detector for each of these effects. Each of the detector would raise a flag. First flag detected ECG misplacement by searching for starlight lines. They detected straight lines using derivative after splitting the signal into five non-overlapping segments. Second flag detected large impulse. To detect it, the authors segmented the ECG signal into ten non-overlapping segments and set a threshold. If a segment had an amplitude the exceeded the threshold then they counted it as an impulse. Another flag detected the existence of Gaussian noise. Detecting Gaussian noise was based on using power spectrum with AR model accompanied with sample entropy. The ratio of noise (signal with frequencies > 40Hz) to intrinsic signal (signal with frequencies between 0.05-40Hz) and the sample entropy decided whether to raise this flag. The last flag recognized QRS waveform detection error. The R peaks were detected and the duration between R peaks were considered as heartbeats. Impulse
rejection filter was used to detect errors in heartbeats duration. From all these flags, a quantitative assessment was reported.

In [95], seven simple measurements were analyzed to report an ECG signal quality measure. Six of the measurements were frequency components of ECG signal while the seventh measurement was for ECG signal electrodes movements. The authors based their work assuming ECG signal spanned frequencies from 0.05-100Hz. The frequency bands were as the following: 0.05-0.25Hz was considered as the low frequency, lower ECG bandwidth occupied the bandwidth from 0.25-10Hz while higher ECG bandwidth occupied frequencies of 10-20Hz. The medium frequency range spanned the frequencies from 20-48Hz. The power-line frequency was captured in the range from 48-52Hz. This was because the authors used a dataset from a country that had 50Hz power-line, and the last frequency band was the high frequency which occupied the range from 52-100Hz. The seventh measurement captured electrode movement by assigning a $\pm 4\text{mV}$ amplitude threshold. These bands were compared on several subjects and on different times to measure quality of an ECG signal.

In [96], ECG heartbeat were extracted as the duration of $[-200:400]$ milliseconds around the R peak. Two methods were proposed for outlier eliminations: one method used the mean (or median) of each subjects heartbeats and calculated a template. Within that template, the maximum and the minimum R peaks and distances for the heartbeats that constructed a template was recorded. Distance measure between every heartbeat and its template was based on Euclidean or cosine distance. Based on the distance and comparison to template parameters such R peaks and minimum and maximum distances, the signal was classified as normal or abnormal. The other method applied clustering with the concept of density reachability. Every heartbeat was compared to the densities, and if it was within the preset threshold distance from the density and had a preset number of neighbors then it was considered as a normal heartbeat; otherwise, it was classified as an outlier.
The aforementioned techniques dealt with ECG signals whether they were used for biometrics purposes or medical diagnoses applications; however, the area of outlier detection has been widely examined in statistics, data mining, sensor networks, and other areas. A detailed survey is in [97]. The work in [98] studied outlier detection for temporal data using several methods, including but not limited to, AR models, Markov models, histograms, neural network, distance and density approaches, clustering, and other approaches.

Most, if not all, of the reviewed techniques are offline. They are offline in a sense that they operate after collecting the signal. In other words, they require future observations to make a decision. In this thesis, we desired an online outlier removal technique. Table 2.2 summarizes the reviewed literature on irregular ECG heartbeat detection and removal.
<table>
<thead>
<tr>
<th>Citation</th>
<th>Method</th>
<th># sub.</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gokana et al. [61]</td>
<td>RMSSD and correlation</td>
<td>47</td>
<td>99.5%</td>
</tr>
<tr>
<td>Ayub et al. [64]</td>
<td>correlation</td>
<td>47</td>
<td>98.56%</td>
</tr>
<tr>
<td>Maharaj et al. [31]</td>
<td>wavelet correlation and variance</td>
<td>148</td>
<td>FRR 95%-96% and FAR of 92%-96%</td>
</tr>
<tr>
<td>Mašetic et al. [65]</td>
<td>auto-regression Burg algorithm and decision tree</td>
<td>15</td>
<td>99.86%</td>
</tr>
<tr>
<td>Yu et al. [66]</td>
<td>morphology, energy, and coherence of wavelets as features and neural network classifier</td>
<td>23</td>
<td>99.65%</td>
</tr>
<tr>
<td>Mohamad et al. [69]</td>
<td>fiducial points were extracted from three leads as features and Elman neural network was used for classification</td>
<td>1,600 segmented heartbeats</td>
<td>≥ 95%</td>
</tr>
<tr>
<td>Hayes et al. [70]</td>
<td>wavelet coefficients were used as features and SVM was used for classification</td>
<td>109,494 heartbeats</td>
<td>≥ 90%</td>
</tr>
<tr>
<td>Authors</td>
<td>Methodology</td>
<td>Number</td>
<td>Accuracy</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Oresko et al. [72]</td>
<td>Fiducial points were used for extraction and feed-forward neural network was used for classification</td>
<td>48</td>
<td>≥ 90%</td>
</tr>
<tr>
<td>Shiu et al. [74]</td>
<td>Used high-pass filter on signals excluding their QRS complex then sliced their ST waveform</td>
<td>*</td>
<td>FRR of 96.6% and FAR of 93.1%</td>
</tr>
<tr>
<td>Yeh et al. [79]</td>
<td>Fiducial feature extraction and LDA was used for classification</td>
<td>48</td>
<td>Average of 94.48%</td>
</tr>
<tr>
<td>Lanata et al. [83]</td>
<td>Cumulant Fourier transform for feature extraction and KNN was used for classification</td>
<td>**</td>
<td>94.33 ± 0.52%</td>
</tr>
<tr>
<td>Prasad et al. [84]</td>
<td>Uses higher order spectra for feature extraction and KNN for classification</td>
<td>2,383</td>
<td>97.65%</td>
</tr>
<tr>
<td>Naseri et al. [93]</td>
<td>Used signal energy to eliminate outliers</td>
<td>1,500</td>
<td>93.6%</td>
</tr>
<tr>
<td>Liu et al. [94]</td>
<td>Operated on 12-lead signals using flags to detect irregularities</td>
<td>998</td>
<td>FRR=90.67% and FAR=89.78%</td>
</tr>
<tr>
<td>Allen et al. [95]</td>
<td>seven measurements on ECG signals</td>
<td>20</td>
<td>***</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>Lourenço et al. [96]</td>
<td>applied mean (median) distance to a template and used clustering with density reachability</td>
<td>63</td>
<td>EER=12.4%</td>
</tr>
<tr>
<td></td>
<td>- used clustering with density reachability</td>
<td></td>
<td>EER=13.4%</td>
</tr>
</tbody>
</table>

Table 2.2: Summary and highlights of irregular ECG heartbeat detection and removal. *mentions the data is from MIT-BIH arrhythmia and European ST-T database. **mentions MIT-BIH arrhythmia database but does not mention if all or parts of the database was used. ***the work is about quality measure of a signal, so no accuracy was reported
2.3 Data Synthesis and Small Sample Size Problem

The problem of having small sample size persists among most biometric systems. There are several approaches in the literature that attempted tackling this issue, and here are some of the techniques that we reviewed and those that helped us in our work in Chapter 5. Small sample size issue occurs when the number of observations is much smaller than the number of variables to model. Several approaches are available to tackle this problem such as dimensionality reduction, data synthesize, and cascade classifiers which deal with data imbalance. Parts of this thesis study the effect of such problem on ECG and propose a viable solution.

Most of the work in the literature apply dimensionality reduction techniques. In [99], the authors claimed that when dimensionality reduction was used, the accuracy increased when sample size increased; however, it decreased when a specific sample size was reached. Feature selection and feature extraction are other approaches to handle having small sample size, and they are similar in concept to dimensionality reduction. The work in [100] acknowledged the dimensionality reduction problem but claimed that using SVM could be a viable approach since SVM generalizes with small sample size and high dimensional space. In [101], the authors examined several dimensionality reduction techniques with different feature selection methods (Wrapper and ReliefF), feature extraction (PCA), and classifiers (KNN, LDA, Naive Bayes, SVM and others). It was demonstrated that the highest accuracy was achieved using ReliefF and PCA since they could better generalize the data. In [102, 103] the authors proposed quadratic-like discriminative analysis.

Generating synthesized data is mostly examined in face recognition due to the symmetry of the face. In [104], the authors generated mirror images of the original face image, and generated extra left and right symmetrical images. In [105], the authors proposed a cascade classifier where each classifier was trained with a fixed number of samples to reduce data imbalance. Lastly, some techniques were oriented towards synthesizing ECG heartbeats as in [106]. These techniques were not designed for biometrics applications.
but for biomedical signal processing applications.

In ECG signal biometrics, most of the works apply one or more types of dimensionality reduction techniques and/or feature selection. We investigated dimensionality reduction techniques in subsequent chapters of this thesis. Our emphasis was oriented towards studying the ECG signal statistics itself and the capability to synthesize it in order to improve biometric system accuracy. Unfortunately, there is very few, if any, ECG signal synthesis techniques that is used to improve ECG biometrics performance.

2.4 Continuous Authentication

There are few papers on CA and ECG biometrics together, and this section reviews the techniques that attempted resolving similar problems as ours, specifically for Chapter 6. The work in [2] first preprocessed the ECG signal by applying a notch filter to remove the power-line interference. A third order high-pass Butterworth filter with 0.5Hz cutoff frequency was applied to reduce baseline wander. The authors used only the QRS complex but for several leads. The QRS waveforms were segmented based on extracting a range of samples around the R peaks. An outlier removal based on autocorrelation coefficient between the segmented QRS and the average QRS was used. When the correlation of QRS waveform was less than a specified threshold, it was removed. On the contrary, the QRS waveform with the highest coefficient was considered as the template. Afterwards, several decision-making strategies were experimented including median, mean, 75\textsuperscript{th} percentile, 90\textsuperscript{th} percentile, 95\textsuperscript{th} percentile, and max to obtain a single classification decision. The segment size, which is the number of samples (observations) needed to make a decision, was predetermined.

In [107], the research was fiducial points based approach that extracted time durations and amplitude differences. It extracted 24 features from ECG heartbeats. Outlier removal based on the extracted features was deployed. It used a range of 1 standard
deviation from the mean as an acceptable range; otherwise, the features were removed. The authors also used a feature selection technique to reduce the number of features from 24 to nine features. Mahalanobis distance was used for classification. In the continuous authentication, 35 heartbeats were used to make a decision. If the number of successes passed a threshold (30 in the paper), the segment was considered as a positive authentication; otherwise, an imposter was decided as taken control. The work presented in [3] used autocorrelation on an ECG signal, once the value dropped to a pre-specified threshold, a request was sent to update the model.

The research paper [108] applied a string matching technique based on Ziv-Merhav (ZM) cross-parsing [109]. This method applied an incremental LZ-type parsing algorithm. In this paper, ECG signals were encoded into an 8-bit uniformly quantized alphabet, and they were classified with KNN with K=1. To implement a continuous authentication system, the authors continuously updated the genuine subject model once a positive authentication occurred. FIFO (first in first out) was used to deal with model size limitation. In this work, every ten heartbeats were grouped together.

To the best of our knowledge, these are the only papers that apply continuous authentication on ECG signals.

Nevertheless, continuous authentication was examined for other biometrics traits, especially keyboard keystroke and mouse movement, among others. The work in [110] examined mouse movement with different strategies and settings including static and dynamic trust model, fusion, and score boosting algorithms. The concept of trust model was that the examined subject was compared to its genuine model, and if the resemblance was less than a predefined threshold, then this subject was locked out. Fusion on the other hand, is a technique that combines features, classifiers confidence values, or classifiers class. Based on the combination, a decision was made on the identity of the subject. The score boosting technique was a method that remapped the classifier score such that it kept it unchanged when it lied between two thresholds (upper and lower) while it
boosted it up when it was above the upper threshold and boost it down when it was below the lower threshold. Throughout experiments, SVM was used for classification.

In [1], mouse movements were examined as well. The features were extracted from distance traveled, time, and direction. The features were histograms of occurrences of actions in a predetermined duration. Neural network was applied for classification, and sequential sampling was used for continuous authentication. Technical details pertinent to this work is explained in Section 7.3.1.

In [111], mouse actions were also used for biometrics and continuous authentication. Some of the actions were direction, clicking patterns, spatial location on the screen, and several other actions captured by the computer. One-class detectors including nearest-neighbor detector, SVM, and neural network were used. The one-class detector modeled the genuine subjects using their data only. Afterwards, it considered subjects that were further away from the model by a threshold as imposters. A 28-subject database was used in the experiment, and one subject in every iteration was used in the model. Ten-fold cross-validation was examined when collecting results. The one-class SVM reported the highest accuracy, and its performance was followed by neural network. Nearest neighbor achieved comparably the lowest results. Segment size, in this case authentication time, was also predetermined, and it showed that the longer the authentication period the higher the accuracy. For instance, when one minute session was examined, a false acceptance rate of 44.56% and a false rejection rate of 34.78% were achieved. On the other hand, when the session was ten minutes, a false acceptance rate of 2.75% and a false rejection rate of 3.39% were achieved. The acceptability of the session length was a trade-off between usability versus security. A ten-minute session might be long enough for the imposter to make the desired damage while a minute session had a performance close to being random.

On the other hand, for keyboard CA systems, the work in [112] was trained by making users type predefined words to generate a model. These words were chosen particularly
by analyzing the frequency of words typed in e-mails for fifteen users. Some of the the
most frequent words were: dear, hi, hello, regards, thanks, meeting, time, AM, and
PM. The purpose of using most frequent words was the hypothesis that subjects typing
patterns are more stable writing a word they frequently use. The feature extraction was
based on time to type a word, and time between typing and releasing keyboard buttons.
Decision measurement was made for every typed word based on the standard deviation
of Euclidean distance. The decision measurement score was compared to predefined
thresholds, and a decision was made accordingly. Furthermore, a work similar to the
work in [110] was applied to keystroke biometrics in [113]. The authors in [113] suggested
a method to reduce the needed sample size by predicting a future desired probability from
a smaller sample size.

Lastly, all the studies presented in this subsection utilized unimodal biometrics. More
works fused different biometric modalities to achieve a continuous authentication system.
In [114], face recognition and clothes color were proposed. Also the work in [115] fused
fingerprint and face recognition.

The literature of continuous authentication either uses multi-modal biometrics or re-
quire a pre-set segment size (sample size or number of observations) and decision thresh-
olds. These techniques cannot be used to achieve the objective of this thesis, which is
a continuous authentication system that does not require decision making strategy, and
does not need a preset segment size and decision thresholds. Table 2.3 summarizes all
the reviewed continuous authentication methods along with types of feature extraction,
types of classifiers, accuracy achieved, and size of database used.
<table>
<thead>
<tr>
<th>Citation</th>
<th>Type of biometrics</th>
<th>Method</th>
<th># subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labati et al. [2]</td>
<td>ECG</td>
<td>several decision making strategies with a preset segment size</td>
<td>184</td>
</tr>
<tr>
<td>Guennoun et al. [107]</td>
<td>ECG</td>
<td>fiducial based approach with a preset segment size</td>
<td>16</td>
</tr>
<tr>
<td>Coutinho et al. [109]</td>
<td>ECG</td>
<td>used incremental LZ-type parsing algorithm and KNN with model update capability</td>
<td>19</td>
</tr>
<tr>
<td>Mondal et al. [110]</td>
<td>mouse movement</td>
<td>used dynamic trust model, fusion, and score boosting. SVM was used for classification.</td>
<td>49</td>
</tr>
<tr>
<td>Ahmed et al. [1]</td>
<td>mouse movement</td>
<td>sequential sampling in a predetermined window size classified using neural network</td>
<td>5</td>
</tr>
<tr>
<td>Shen et al. [111]</td>
<td>mouse movement</td>
<td>one-class SVM achieved the highest accuracy</td>
<td>28</td>
</tr>
<tr>
<td>El Menshawy et al. [112]</td>
<td>keyboard keystroke</td>
<td>was trained on specific words to model typing speed stroke time</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 2.3: Summary and highlights of the reviewed continuous authentication methods
2.5 Chapter Summary

Research in ECG biometrics has first started in the 1970’s, yet most of the contributions have been done after the millennium. ECG as a signal is prone to noise. This chapter described the advancement in ECG biometrics in general and delved into more comprehensive explanations in the area where this thesis contributed the most.

As in Section 2.1, research in ECG biometrics takes two mainstreams: fiducial points based approach and non-fiducial points based approach. In fiducial points based approach, amplitude, durations, angles of fiducial points are used to construct the feature vector. On the other hand, non-fiducial points based approach is a holistic approach that does not mainly rely on fiducial points to extract features. It uses one or more fiducial points to segment and align ECG heartbeats. Sometimes, non-fiducial based approach does not use any fiducial points. Difficulties in localizing fiducial points are the main drawbacks of fiducial points based approach, especially when ECG heartbeat irregularity or noise is presented. Data redundancy, on contrast, is the main disadvantage of non-fiducial points based approaches. As it is explained in later chapters, the main database examined in this thesis is a noisy database, and detecting fiducial points is not the optimum approach. Hence, we followed the non-fiducial points based approach, and we proposed a feature extraction method that extracts temporal variations among observations.

Noise and irregular observations, in this case heartbeats, negatively affect biometric system accuracy. We reviewed methods used to reduce the noise and the number of outlier heartbeats. Section 2.2 explains two methods for outlier removal: one method targets specific type of irregularity, and the other approach models regular (*normal*) heartbeat signal and considers anything that is not regular as irregular heartbeats, and eventually, eliminates them. A consequent problem, which may arise by eliminating a significant number of ECG heartbeats is the small sample size. For this purpose, along with the slow acquisition of ECG signal, data synthesis and methods to deal with small
sample size are explained in Section 2.3.

Research in continuous authentication has been investigated in several areas such as keystroke, mouse movement, face recognition and via multi-modal systems; however, it was researched in a lesser extent in ECG biometrics. A property of ECG signal is being quasi-periodic, and we desired to exploit such characteristic to achieve a continuous authentication system that can improve biometric accuracy and capture intruders who attempt hijacking the biometric system. Section 2.4 presented several approaches in the literature to tackle this issues.
Chapter 3

Experimental Set Up

Several databases and methods of evaluation are available for ECG analyses; however, most of the databases are collected and dedicated for medical purposes. Most databases may not be suitable for biometric experiments as biometric databases need to meet the following properties:

- They contain more than one acquisition session for each subject and are collected within different timestamps. Having this feature allows researcher to examine the robustness of ECG signal biometrics towards variation with respect to time.

- They have a large number of subjects. Larger databases achieve scalability in obtaining less performance variance.

- They are extracted in different environments (i.e. from calm subject, exercising subject, or the subject undergoing other daily activities). ECG heartbeats tend to change their morphology depending on the activities performed, and the change in activity may jeopardize the biometric system performance.

This chapter presents the databases examined in this thesis, method of evaluation to assess the performance, and the experimental set up. Two databases and two different applications were examined to demonstrate that the proposed work is not restricted to
a single application or a database. The applications are medical and biometrics applications. In the medical application, classification of healthy versus non-healthy ECG heartbeats were conducted on the Physikalisch-Technische Bundesanstalt database. For the biometrics application, the proposed work was examined on University of Toronto Database for a verification biometric system. A verification biometric system is a binary pattern recognition problem where the output is to accept or reject a claimed identity. A verification system was examined in particular since it is a required stage in continuous authentication system which needs to verify the identity of a person or to flag the person as an intruder.

3.1 Database

3.1.1 Physikalisch-Technische Bundesanstalt (PTB) Database

A publicly available database called the PTB Diagnostic ECG Database [116, 117] was used to assess the performance of parts of the proposed contributions in this thesis. It was used to investigate the viability of some of the contributions on a medical application. This database was collected from healthy and non-healthy persons and is usually used for medical diagnosis. The non-healthy subjects suffer from different heart related diseases. Despite there exist several databases publicly available in [116], PTB was chosen in particular due to having many (290) subjects along with different heartbeats irregularities diseases.

This database was recorded using 12-lead ECG configuration. Three Frank leads were used as well. The 12-lead ECG configuration is illustrated in Figure 1.2. The Frank leads are additional 3 leads added on the back of the subject. This configuration was proposed by Frank Wilson in 1932 [118].

Each ECG signal is sampled at 1KHz with a 16 bit resolution. This database has 549 records from 290 different subjects. All subjects but 22 have summary files that contain
the subject's information such as diagnosis, gender, age, and other information. Each recording has a duration of 2-3 minutes. Data from lead I only were used throughout all experiments. The cases in the database consisted of the following diseases: myocardial infarction, cardiomyopathy/heart failure, bundle branch block, dysrhythmia, myocardial hypertrophy, valvular heart disease, myocarditis, healthy, and other miscellaneous ones.

The database has 52 healthy subjects. Out of the 52 subjects, 14 subjects have more than one recording. Some of these multiple recordings are several years apart.

3.1.2 University of Toronto Database (UofTDB)

This database was collected at the University of Toronto [6]. It was collected in six sessions that spanned six month period. Not all subjects were involved in all six sessions. Different postures were included in the data collection: sit, stand, supine, and tripod. Exercise was also included for some subjects where the elevated heart rate was 132 beat/minute (bpm). Unlike PTB database, UofTDB was collected from fingertips with single lead. The sampling rate for the signal is 200Hz. Out of the 1,020 subjects, 1,012 subjects had come in the first session, 82 subjects had come twice, and out of the 82, 43 had come for the six sessions. There are 71 subjects out of the 1,020 subjects who had exercise collection. In this research, we considered single and multiple sessions subjects, 1,012 and 82 subjects, respectively. Figure 3.1 shows the acquisition equipment and set up for UofTDB database.

Figure 3.1: 1-lead UofTDB acquisition configuration [6]


3.2 Experimental Set Up

The signal from the acquisition device is continuous as in Figure 3.2

![Figure 3.2: Continuous ECG signal for 5 second duration](image)

Classifiers are trained with observations, and an observation, in this thesis, is an isolated heartbeat. Each ECG signal was segmented into observations.

3.2.1 Preprocessing

Some of the main sources of noise in a collected ECG signal include baseline drift, electromyogram signal interference, power-line interference, and contact noise. More details about noise is presented in Section 1.4. Regularly, ECG signal spans the frequencies from either 0.5Hz or from 1Hz up to 40Hz [5, 55]. Some works claim that ECG has intrinsic signal up to 100Hz [20]. Similar to [5, 55], a fourth order band-pass Butterworth filter was applied with cutoff frequencies once of 0.5-40Hz, and another time with frequencies
of 1-40Hz. Results were almost identical whether 0.5Hz or 1Hz was used. With this
bandpass filter, the baseline wander and the higher frequencies of power-line interference
were reduced. Figure 3.3 illustrates the preprocessing effect on a signal.

Figure 3.3: Illustration of ECG signal preprocessing stage
3.2.2 Segmentation

A sub-database was collected for each of the databases. In the PTB database, QRS detection method (GQRS [119]) for segmentation was used while Pan-Tompkins procedure [62] was used on UofTDB. The R peak from the detected QRS waveforms were used to align ECG heartbeats. One second was the segment duration with the R peak being at the center. Several works, including [120, 121, 122], aligned ECG heartbeats based on R peaks. A sample of the aligned sub-database is shown in Figure 3.4.

![3D sample of aligned ECG waveforms](image1)

![Overlapped sample of aligned ECG waveforms](image2)

(a) 3D sample of multiple ECG heartbeats  (b) Overlapped ECG heartbeats

Figure 3.4: Illustration of 67 healthy ECG heartbeats for different subjects

3.2.3 Classification

There are several classification methods in the literature, and bagging [123] is one of them. In a nutshell, bagging is a machine learning technique that generates weak classifiers/predictors. The aggregated average of weak classifiers makes a decision. Bagging was heavily used in this thesis in particular because we observed an unstable classifier prediction when we examined ECG heartbeats data. It was unstable in a sense that slight change in the training data led to a significant change in the construction of the classifier and a significant change in accuracy. Bagging usually reduces this issue [123]. Compara-
son to classifiers, SVM [124] and single decision tree, are presented in Section 6.4.1.

Suppose a training dataset, \( \mathcal{L} \), is populated with data \( \{y_n, x_n, n = 1, \ldots, N\} \), where \( y \) is the data class and \( x \) is the input data. From these samples, bagging generates multiple bootstrap samples, \( \mathcal{L}^{(B)} \), from \( \mathcal{L} \). For each \( \mathcal{L}^{(B)} \), it finds a predictor, \( \phi(x, \mathcal{L}^{(B)}) \), that predicts the class, \( y \). Bootstrapping samples, \( \mathcal{L}^{(B)} \), are constructed by drawing \( N \) samples with replacement from \( \mathcal{L} \). The predictor used with bagging in this thesis is simple decision tree. The final decision on the class is made by voting.

Increasing the number of bootstraps trials might increase accuracy, yet it increases the complexity of the system. Throughout this thesis, we conducted experiments and we find that 50 bootstraps are adequate for the task as in Figure 3.5.

![Bagging algorithm error vs number of decision tree](image)

**Figure 3.5: Bagging with 50 decision trees**

For each re-sampled set of observations, binary decision tree was used. Classification decision tree recursively splits based on the feature that can partition the data with minimum classification cost [125]. CART algorithm [126] was used to train the decision
tree. CART algorithm uses Gini index for its impurity function, and uses cross-validation to measure its error. Gini index for impurity is measured as:

$$GI = 1 - \sum_j p_j^2$$

(3.1)

where $j$ is the number of classes, which is two in this thesis. Figure 3.6 illustrates the training and testing procedure for the used bagging classifier.

### 3.2.4 Medical Application Experiment Set Up

Regular and irregular heartbeats are to be detected; hence, the PTB segmented dataset was divided into regular and irregular datasets. The regular heartbeats corresponded to healthy heartbeats while irregular heartbeats corresponded to non-healthy waveforms.

Randomly selected non-overlapping regular/irregular heartbeats were used to train and test a classifier. A set of 1,000 regular heartbeats and another 1,000 irregular heartbeats were used for training while all other heartbeats in the PTB database were used for testing.

### 3.2.5 Biometric System Experiment Set Up

Single session dataset of 1,012 subjects and multiple sessions dataset of 82 subjects from the UofTDB were experimented in a biometric system. We constructed the training and testing datasets as follows: in the multiple sessions experiment, one session from each subject was (were) used as genuine\(^1\) training dataset and the other session(s) for that subject was used as genuine testing dataset(s). This was repeated for every session and every subject. In other words, every session for every subject was once treated as genuine training dataset and in other times as genuine testing dataset. The other 81 multiple

---

\(^1\)Genuine dataset corresponds to observations from the subject of interest that the verification biometrics should decide on.
sessions subjects formed the imposter\textsuperscript{2} testing dataset. The imposter training dataset that was used to train the classifier was randomly constructed by drawing heartbeats from subjects that did not belong to the 81 multiple sessions subjects.

In single session experiments, the first 80\% of each subject’s heartbeats were used as genuine training dataset while the other 20\% were used as genuine testing dataset. Subjects randomly drawn from the other 1,011 subjects constructed imposter training dataset. The other subjects that were not part of the imposter training dataset created the imposter testing dataset.

### 3.3 Performance Measure

Performance rate (\(PR\)), false acceptance rate (\(FAR\)), false rejection rate (\(FRR\)), number of genuine observations (\(GO\)), number of intruder observations (\(IO\)), receiver operating characteristic (\(ROC\)) curve, and equal error rate (\(EER\)), were the main measures used in this thesis. Each examined dataset has \(G + I\) samples, with \(G\) is being the number of genuine samples, and \(I\) is being the number of intruder or imposter samples. Depending on the examined experiment, if the experiment is for a biometrics application, then \(G\) stands for genuine subjects samples; however, if it is designed for a medical application, then \(G\) stands for regular heartbeats samples. Similar convention goes for \(I\) as being intruder and irregular samples for biometrics and medical applications, respectively.

We define number of true positive, \(nTP\), as the number of correctly classified genuine samples. Similarly, the number of true negative, \(nTN\), is defined as the number of correctly classified imposter samples. Moreover, number of false positive, \(nFP\), is the number of misclassified imposter samples as genuine samples. Likewise, the number of false negative, \(nFN\), is the number of misclassified genuine samples as imposter samples.

Following these definitions,

\textsuperscript{2}Imposter dataset corresponds to observations from any subject that is not the genuine subject.
\[
\text{FAR} = \frac{nFP}{I}, \quad \text{FRR} = 1 - \frac{nTP}{G}, \quad \text{PR} = \frac{nTP + nTN}{G + I},
\]  

(3.2)

\[
\text{FRR} = 1 - \text{TAR}
\]  

(3.3)

Despite the fact that \( PR \) might seem the most appealing quantity since it includes all measures, all quantities in Equation (3.2) should be considered as \( PR \) would be misleading when \( I \gg G \) or \( I \ll G \).

\( GO \) is the average number of observations needed to reach a decision in a continuous authentication system’s segment\(^3\) when a genuine person is examined. On the other hand, \( IO \) is the average number of observations to reach a decision in a continuous authentication system segment when an imposter person is examined. The mean of these numbers are reported. Quantities similar to \( GO \) and \( IO \) were used and reported in [110].

\( ROC \) curve measures the performance of a system in different operating points. \( ROC \) curve plots \( FRR \) versus \( FAR \). The operating points are threshold values that change \( FRR \) from 0 to 1, similarly, change \( FAR \) from 1 to 0. The choice of operating point is an application dependent since there is a trade-off between \( FAR \) and \( FRR \). Closely related to \( ROC \) is \( EER \). \( EER \) is the error on the operating point on which \( FAR \) is equal to \( FRR \).

Lastly, the terms \textit{training error} and \textit{testing error} are usually used in machine learning literature. The term \textit{training error} is called on the error resulted when examining classifiers on training dataset. However, \textit{testing error} is called on the error resulted when examining classifiers on testing dataset. The error is measured as

\[
\text{error} = 1 - PR
\]

\(^3\text{Segment size corresponds to the number of observations needed to make a decision.}\)
3.4 Summary

This chapter is intended as a guide to common details for most experiments in this thesis. The chapter started by highlighting desired features in a database in order to be considered for biometrics applications and to report meaningful and practical results. Further, contributions in this thesis was used for medical application for health assessment. Two databases were used in this dissertation, and they are explained in Section 3.1. One database was used for medical application, PTB database, and the other database was used for biometrics applications, UofTDB.

Throughout all experiments, we started with raw ECG signals, which were continuous signals. We segmented the signal into individual heartbeats as explained in Section 3.2.2. These heartbeats, depending on their application, were split into training and testing datasets. Section 3.2.4 and Section 3.2.5 explain the procedure followed to construct such datasets. Lastly, data reporting and the method to evaluate the experiments are explained in Section 3.3.
Chapter 3. Experimental Set Up

Feature vector 1
Feature vector 2
Feature vector 3
Feature vector N

1st bootstrap. Randomly, and with replacement, draw N feature vectors

Grow a decision tree

2nd bootstrap. Randomly, and with replacement, draw N feature vectors

Grow a decision tree

kth bootstrap. Randomly, and with replacement, draw N feature vectors

Grow a decision tree

Training

Testing

Figure 3.6: Bagging classifier scheme
Chapter 4

Online Outlier Removal

Electrocardiogram signal is prone to noise interference. If noisy signals are not filtered or corrected when used in an automated system such as biometric systems, the performance of the system may degrade. In this chapter, we develop an online abnormal electrocardiogram heartbeat detection and removal. The technique eliminates outliers in real-time while data are being collected. We used one-class classifier concept with Gaussian mixture model to model normal electrocardiogram heartbeat.

4.1 Introduction

ECG signal is a low amplitude signal generated from an internal organ. As explained in Section 1.4, ECG signal is prone to noise interference. Using the 12-lead configuration for biometric systems maybe commercially appropriate to only some applications such as tele-medicine where ECG signal is collected for medical purposes. However, using 12-lead configuration as a wearable device for daily purposes may not be very attractive due to its electrodes set up inconvenience. Other configurations, such as a 1-lead configuration [6] that collects ECG signals from fingertips using three electrodes is more appealing. Such configuration is more prone to noise than the 12-lead configuration.

In ECG signal biometrics, an observation is passed to a classification module to make
a decision to accept or reject an identity, in the case of a verification system or finding
the identity of the person, in case of an identification system. Passing a noisy or a
deformed signal may jeopardize the classifier decision. We develop an online and real-
time abnormal electrocardiogram heartbeat detection and removal. In this chapter, the
term online indicates that the system operates without the need for the entire signal to be
collected first. Online systems depend on current and previous observations only. Using
online outlier removal, the system is capable to provide an instant indication about the
“goodness” of the collected signal. We used Gaussian mixture model to model normal
electrocardiogram heartbeat. This outlier removal method was utilized in a biometric
system and was examined on a fingertip acquired ECG signals database. Comparison to
other outlier removal systems shows the superiority of the proposed method.

4.2 Methodology

In this section we explain the developed method that achieved an online ECG signal
outlier detection for biometrics purposes. The method consists of three main stages: first,
a filtration stage was applied to reduce the low and high frequency noise in the ECG quasi-
periodic signal. Afterwards, QRS waveform detection was utilized to segment ECG signal
into individual heartbeats. Lastly, Gaussian mixture model (GMM) was used to model
normal heartbeats. This model discarded heartbeats that were statistically different than
the trained model. The first two stages are explained in Section 3.2.1 and Section 3.2.2,
respectively. The GMM method is explained in this chapter.

4.2.1 Gaussian Mixture Model (GMM)

GMM is usually used as an unsupervised clustering method. However, in this thesis
it is used to model normal ECG heartbeats. GMM is a sum of M weighted Gaussian
densities [127] given by
Chapter 4. Online Outlier Removal

\[ P(\mathbf{x}) = \sum_{m} w_m p(\mathbf{x}, \mu_m, C_m) \quad (4.1) \]

where \( w_m \) are the weights of the Gaussian densities, \( \sum_{m} w_m = 1 \). \( \mathbf{x} \) is a \( k \) dimensional features vector. Therefore, the pdf, \( p(\mathbf{x}, \mu_m, C_m) \), is

\[ p(\mathbf{x}, \mu_m, C_m) = \frac{1}{(2\pi)^{\frac{k}{2}} |C_m|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x} - \mu_m)^T C_m^{-1}(\mathbf{x} - \mu_m)} \quad (4.2) \]

where \( \mu_m, C_m \) are the mean vector and the covariance matrix, respectively. Also, \( |C_m| \) is the determinant of the covariance matrix.

The motivation behind using the GMM was the assumption that normal ECG heartbeats could be modeled into \( M \) Gaussian densities, each in \( k \) dimensions.

Expectation maximization (EM) [128] algorithm was used to construct the GMM. EM considers all training examples and tries to fit a Gaussian distribution on it. The training algorithm would be as the following:

1. Compute the probability that the training \( \mathbf{x} \) belongs to the Gaussian \( m \) using
   \[ P(\mathbf{x}|m) = \frac{w_m^{(i)} p(\mathbf{x}|\mu_m^{(i)}, C_m^{(i)})}{\sum_{j} w_j^{(i)} p(\mathbf{x}|\mu_j^{(i)}, C_j^{(i)})} \] Where \( p(\mathbf{x}|\mu_m^{(i)}, C_m^{(i)}) \) is used to indicate that these values depend on the previous iteration

2. Estimate the new weight \( w_m^{(i+1)} = \frac{1}{T} \sum_{t=1}^{T} P(\mathbf{x}_t|m) \)

3. Estimate the new mean \( \mu_m^{(i+1)} = \frac{\sum_{t=1}^{T} P(\mathbf{x}_t|m) \mathbf{x}_t}{\sum_{t=1}^{T} P(\mathbf{x}_t|m)} \)

4. Update the variance \( \sigma_m^{2(i+1)} = \frac{\sum_{t=1}^{T} P(\mathbf{x}_t|m) (\mathbf{x}_t - \mu_m^{(i+1)})^2}{\sum_{t=1}^{T} P(\mathbf{x}_t|m)} \)

Where \( T \) is the number of observations in the training dataset. There is no specific method for termination; however, it is usually based on a heuristic approach.

**Evaluation procedure**

After obtaining the Gaussian models from the training data, the evaluation was based on the negative minimum log-likelihood measure. The minimum log-likelihood measures quantitatively the likelihood that the tested data belongs to a mixture.
The lower negative log-likelihood the more similar the data to be of the model. Therefore, choosing the minimum of the negative log-likelihood is equivalent to choosing the maximum likelihood.

### 4.2.2 Modeling Normal Heartbeats

It is desired to achieve a real-time outlier removal system. We approached this problem by creating a generic model that is capable to distinguish normal from abnormal heartbeats. The model is subjects invariant; hence, it can be used in biometric system without requiring subjects’ specific data. In this section, we study the feasibility of using GMM to detect abnormal heartbeats. Also, we examine the influence of different number of Gaussian mixtures.

If we model normal heartbeats then any heartbeat with statistics significantly different from the normal heartbeat model is classified as an abnormal heartbeat. Hence, we constructed a normal heartbeats model. For the task, normal heartbeats segments were collected to train the GMM.

In this chapter, single session database of UofTDB in Section 3.1.2 was used. The preprocessing stages explained in Section 3.2.1 was applied to each subject’s ECG signal. From the 1,012 subjects, a total of 158,984 heartbeats segments were available for experimentation.

From the database of 158,984 heartbeats, 1,330 heartbeats were separated into normal and abnormal heartbeats. Out of the 1,330 heartbeats, 930 heartbeats were normal heartbeats while the other 400 heartbeats were abnormal heartbeats. It is worth mentioning that the subjects whose normal and abnormal heartbeats were used in constructing the GMM model were removed from the database in the experiments. As a result, the experimentation database consisted of 943 subjects.

The criteria to separate normal and abnormal heartbeats was based on achieving two purposes: first, we needed to eliminate the obviously abnormal heartbeats, and second,
we should allow interclass and intraclass variations. Consequently, the strategy we followed was that a heartbeat was considered as *abnormal* if it visually looked significantly different than the ECG heartbeat in Figure 1.3. Figure 4.1 illustrates an example of a *normal* and an *abnormal* ECG heartbeats.

Out of the 930 *normal* heartbeats, 800 heartbeats were used to train a GMM model while the other heartbeats were used for model validation. Figure 4.2 presents *EER* for different models.

It can be noticed from Figure 4.2 that when *normal* data were modelled with two Gaussian mixtures, it achieved the best results. All available 930 *normal* heartbeats were used to train a new Gaussian model with 2 mixtures for the experiments. An example of a subject’s heartbeats before and after applying outlier removal using GMM with two components (GMM, M=2) can be observed from Figure 4.3 where 46 heartbeats out of the 125 heartbeats were removed.
4.3 Experimentation

We suggested a method to achieve a heartbeat outlier elimination by modelling normal heartbeats with density estimation using GMM statistics. In this section we examined the viability of such approach. This experiment intended to demonstrate the influence of outlier removal on a verification biometric system. Three biometric systems were implemented and their ROC curves were compared. One of the biometric systems utilized the GMM, $M=2$ outlier removal. Another biometric system had interquartile range (IQR) outlier removal, and the third biometric system was based on a state-of-the-art, AC/LDA [3], which is a biometric system that uses autocorrelation threshold for outlier removal.

Bagging, as explained in Section 3.2.3, was used as a classification stage in GMM,
Chapter 4. Online Outlier Removal

Figure 4.3: GMM, M=2 model outlier removal

M=2 and in IQR based systems. From the 943 subjects, two non-overlapping databases were constructed: training and testing. Following Section 3.2.5, the training dataset consisted of 80% of the data of each subject while the other 20% constructed the testing dataset.

The IQR outlier [129] removal criterion is a statistical method that removed any heartbeat with R peak amplitude greater than $U_{\text{outlier}}$ or less than $L_{\text{outlier}}$ where:

$$U_{\text{outlier}} = Q_3 + 1.5 \times IQR$$
$$L_{\text{outlier}} = Q_1 - 1.5 \times IQR$$

(4.3)

$Q_1$ and $Q_3$ are the 25$^{th}$ and 75$^{th}$ quantiles, respectively. Also, $IQR = Q_3 - Q_1$. R peaks amplitude greater than $U_{\text{outlier}}$ or less than $L_{\text{outlier}}$ were considered outliers and were removed. Figure 4.4 illustrates the boxplot for the R peak amplitude for five subjects. Figure 4.5 shows a subject’s heartbeats before and after outliers removal. Both Figure 4.4 and Figure 4.5 are offline IQR outlier removal.

We implemented two types of IQR outlier removal: online and offline implementations. The difference between them is that the former had the parameters $U_{\text{outlier}}$ and $L_{\text{outlier}}$ pre-set by calculating their values from the 1,330 heartbeats. Hence, similar to the
GMM approach, it can be used to remove outliers in real-time and for any subject. On the other hand, the offline implementation calculated $U_{\text{outlier}}$ and $L_{\text{outlier}}$ for each subject after collecting all the subject’s data. After collecting the parameters, the offline IQR outlier removal re-examined the heartbeats using these parameters and decided whether a heartbeat is a normal or an abnormal heartbeat. Therefore, this latter system is not suitable for real-time applications when unseen-before subjects are encountered.

We compared the implemented verification system to a state-of-the-art-biometric verification system, AC/LDA [3]. AC/LDA system does not require signal segmentation, and it uses autocorrelation as features. It then applies LDA. Lastly, it uses Euclidean distance for classification. It is pertinent to mention that AC/LDA has an outlier removal mechanism based on the Euclidean distance between the autocorrelation values of an observation to the median autocorrelation values for all observations for that subject. It

**Figure 4.4:** Box plot for 5 subjects showing IQR details
decides on the existence of an outlier based on a preset distance value. Hence, following the definition we used of online versus offline outlier removal, AC/LDA is considered as an offline approach. ROC curves for the biometric systems are presented in Figure 4.6. From Figure 4.6 and Table 4.1, improvement is apparent in GMM, M=2 in comparison to other biometric systems. When IQR offline outlier removal was implemented as an online approach, its performance deteriorated due to the large inter-subjects variation in R peak amplitudes.

It is worth noticing the number of observations that became available to the biometric system after applying the outlier removal system in Table 4.1. With GMM, M=2, almost 50% of the original data were classified as irregular heartbeats and were removed.

### 4.4 Summary

Gaussian mixture model (GMM) provided a viable approach to successfully classify normal and abnormal heartbeats. It was used as ECG outlier removal to reduce the number of abnormal ECG heartbeats that may otherwise sabotage biometric system decision if not eliminated. This outlier removal method does not require prior knowledge about the
Table 4.1: EER for the biometric systems. Number of observations corresponds to the number of observations available to the biometric system after applying the outlier removal step. *AC/LDA does not have number of observations since it does not segment the ECG signal.

<table>
<thead>
<tr>
<th>Method</th>
<th>EER (%)</th>
<th>No. of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC/LDA [3]</td>
<td>12.30</td>
<td>*</td>
</tr>
<tr>
<td>No outlier removal</td>
<td>9.44</td>
<td>158,984</td>
</tr>
<tr>
<td>IQR [129]: online</td>
<td>8.25</td>
<td>126,777</td>
</tr>
<tr>
<td>IQR [129]: offline</td>
<td>6.35</td>
<td>143,470</td>
</tr>
<tr>
<td>GMM, M=2</td>
<td>5.94</td>
<td>78,655</td>
</tr>
</tbody>
</table>

Figure 4.6: ROC curves for the biometric systems
tested subject, it can be implemented as an online outlier removal within ECG acquisition devices, and it can perform in real-time. Additionally, this technique provides an instant feedback about the quality of the signal being collected.

In this chapter, normal ECG heartbeats were modeled with GMM. The normal heartbeats segments were collected manually from a database of 1,330 heartbeats. The collection was conducted by removing abnormal ECG heartbeats from the examined pool of heartbeats. A heartbeat that significantly differed from healthy ECG morphology, which contains P, Q, R, S, T, and U fiducial points, was considered as an abnormal heartbeat. In other words, the R peak of the heartbeats were first detected by Pan-Tompkins algorithm then these heartbeats were manually inspected to ensure they follow the morphology in Figure 1.3 to decide whether they are normal or abnormal heartbeats. GMM with two mixtures achieved an EER of $\approx 12\%$ in classifying normal and abnormal heartbeats.

The biometric system that used GMM with two mixtures for outlier removal outperformed the AC/LDA state-of-the-art biometric system by achieving 5.94% EER in comparison to 12.30% EER for the latter. Also, when biometric system with GMM of two mixtures was compared to a biometric system that used IQR online outlier removal, the latter achieved an EER of 8.25%.

Apart from achieving a promising results in constructing an online outlier removal system, this phase has sparked interest in studying the Gaussianity of ECG heartbeats.
Chapter 5

ECG Heartbeats Gaussianity and Data Synthesis

Electrocardiogram is a slow signal to acquire, and it is prone to noise. It can be inconvenient to collect large number of clean ECG heartbeats in order to train a reliable biometric system. In this chapter we study ECG heartbeats Gaussianity, and we generate synthesized data to increase the number of observations. Data synthesis in this thesis is based on the assumption that ECG heartbeats exhibit a multivariate normal distribution while daily ECG variations deviate the signal from Gaussianity; therefore, one can generate ECG heartbeats from such distribution. Lastly, we stabilize performance accuracy by fusing several classifiers in a parallel scheme. This Chapter appears in [130].

5.1 Introduction

Biometric systems require a training stage (interchangeably called enrollment stage) to verify/identify individuals. During the training stage, subjects identities are modeled and stored in a database. Intuitively, the bigger the sample size (the number of observations) the better the model. However, collecting large number of subjects’ training data can sometimes be troublesome. For example, in forensic applications one may have few
fingerprints or mugshots of a subject to model. Collecting training data can also be expensive and inconvenient. For instance, ECG signal would require minutes of clean data acquisition to construct a distinctive dataset. Sparing such amount of time might not be feasible. Airport is a fast-paced environment example where requiring minutes for training data acquisition might not be feasible.

In this chapter we study the Gaussianity of ECG heartbeats and propose an approach to the problem of having small sample size. We have small sample size phenomena when the number of observations is much smaller than the number of variables to model. The time-samples of ECG heartbeats are the variables. Several research papers have tackled such problem [131, 132] especially in face recognition and fingerprint research; however, the amount of details in a single fingerprint or a face image is significantly more than that in an ECG heartbeat. So it can be anticipated that finding distinctive features in ECG biometrics from small number of samples to be more challenging than many other biometrics traits.

There are mainly two issues that may lead to the problem of having small sample size in the context of this thesis. First, if the signal is noisy; thus, removing noisy heartbeats reduce the number of observations in the dataset. Table 4.1 is an example where we notice that \( \approx 50\% \) of ECG heartbeats were eliminated when GMM, \( M=2 \) was applied. Also, the ECG heartbeat is a slow signal to acquire especially if compared to other biometrics traits such as video based face recognition where it is possible to stream 30 frames per second. For practical application, the extent of people’s patience to cooperate and provide their data for enrollment has been recently studied [133, 134, 135].

In this chapter, we simulate the environment of having a limited number of observations for training to imitate having a small sample size phenomenon. We chose an arbitrary number of 20 observations as our baseline since we have an aim to have 30-40 seconds of an enrollment session. This length of ECG signal acquisition can provide a possible range of 20-30 clean heartbeats. We arrived at this number based on observations
from experiments in Chapter 4.

5.2 Methodology

Verification biometric system is the focus of this chapter. As mentioned in Chapter 1, verification biometric system is a binary classification problem to separate two classes: genuine and imposter. The genuine class corresponds to data acquired from the subject that needs to be modeled while the imposter class corresponds to data collected from subjects other than the genuine subject. It is usually the case that imposter class dataset is much larger than the genuine class dataset. In classification problems, classifiers need to be trained with both genuine and imposter datasets in order to design a function that can distinguish between them. Accuracy might be sacrificed if an imbalanced number of data is used. If the number of imposter data is reduced to be in balance with the number of genuine data, the biometric system might not perform as well as if the number of genuine data is increased to be in balance with the number imposter data. Table 5.1 presents this phenomena.

5.2.1 ECG Heartbeats Synthesis

We hypothesize that ECG heartbeats exhibit a multivariate Gaussian distribution. However, the influence of internal and external factors deviate the model from Gaussianity. We attempt to capture this underlying Gaussianity. Each observation consisted of 200 time-samples (random variables) because the sampling rate for UofTDB is 200Hz, and we segmented each heartbeat to have 1 second duration as explained in Section 3.2.2. As mentioned earlier and as shown in Table 5.1, we desired to generate data that can be appended to the genuine dataset to reduce data imbalance and to reduce bias towards imposter dataset.

We modeled the genuine data $X \in \mathbb{R}^{n \times k}$, where $n$ is the number of observations and
Table 5.1: Experiment illustrating data imbalance influence on accuracy. We used 20 observations for genuine set. Despite the fact that EER was not influenced greatly when number of imposter data increased, TAR has decreased significantly and TRR has increased. This suggests that the classifier became biased towards imposter data. EER, TAR, and TRR quantities and their calculations are explained in Section 3.3. TRR and TAR are calculated for the 50% decision threshold of selection between imposter and genuine classes using bagging that is explained in Section 3.2.3.

\[ k = 200 \] is the number of dimensions. Therefore, an observation \( x \) with \( k \) dimensions has probability density \( p(x) \sim N(\mu, \Sigma) \) such that:

\[
p(x) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1} (x-\mu)}
\]

(5.1)

Where \( \mu \in \mathbb{R}^{1 \times k} \) is the mean of \( X \), and \( \Sigma \) is the covariance matrix of \( X \), \( |\Sigma| \) is the determinant of the covariance matrix, and \( \Sigma^{-1} \) is the inverse of the covariance matrix. A synthesis observation is generated by drawing a random vector from this distribution.

The results of data synthesis in Figure 5.1 was not surprising. Prior to making such multivariate assumption, we analyzed the Gaussianity of ECG heartbeat. Analyses have suggested us to model ECG heartbeats into \( k \) multivariate normal distribution. We used Royston’s test [136, 137] for multivariate normality test. It is based on Shapiro-Wilk’s test [138], a univariate normality test. Royston’s test checks normality of each variable alone using Shapiro-Wilk’s test, then it combines Shapiro-Wilk’s statistics into
Figure 5.1: Synthesized data generation from multivariate Gaussian distribution
one statistics test for multivariate distribution. The combined multiple statistics would approximate a chi-squared, $\chi^2$, random variable when the data is a multivariate Gaussian distribution. If $W_j$ is the Shapiro-Wilk’s test of the $j^{th}$ variable in the multivariate data, then Royston’s test, $R$ [139, 140]:

$$ R_j = \left[ \phi^{-1}\left( \frac{1}{2} \phi^{-1}\left( \frac{1 - W_j}{s} \right) \right) \right]^2 $$  \hspace{1cm} (5.2)

Parameters $g, m, s$ are calculated from polynomial approximation. $\phi(\cdot), \phi^{-1}(\cdot)$ are the CDF and its inverse for the Gaussian distribution, respectively. If we have $p$ variates, then the aggregation of $R_j$ in equation 5.3 would have a $\chi^2$ distribution.

$$ H = e \sum_{j=1}^{p} \frac{R_j}{p} $$  \hspace{1cm} (5.3)

$e$ is the equivalent degree of freedom and is calculated as:

$$ e = \frac{p}{1 + (p - 1)C} $$  \hspace{1cm} (5.4)

Where $C$ is calculated as the average of the correlations of $R_j$’s. Furthermore, we utilized Sequential Forward Selection (SFS) [141] algorithm with Royston’s test on training dataset to investigate the number of variables that can constitute a multivariate normal distribution. SFS with multivariate normality tests operates as in Algorithm 1.

After running Algorithm 1, ECG heartbeats could successfully have multivariate normality with more than 20 variables out of the 200 variables. In other words, around 20 out of 200 dimensions could constitute a multivariate normal distribution. This multivariate Gaussianity helps us capture the underlying Gaussianity of the heartbeats and supports our hypothesis that it is most likely that ECG heartbeats exhibit a multivariate Gaussian distribution if there is no changing factors that affect its morphology. Also, experiments based on such assumption improved biometric system performance.

In other words, we assume that ECG heartbeats for each individual exhibit a multivariate Gaussian distribution; nevertheless, the changes in ECG heartbeats morphology
Algorithm 1: SFS for multivariate normality selection

Data: $X \in \mathbb{R}^{n \times k}, X = \{x_1, x_2, \ldots, x_k\}$

Result: $\{S : S \subseteq X, \text{ and is multivariate normal}\}$

Initialization:
An empty set, $S = \{\}$

while (1) do
    while $i \leq k$ do
        if $i=1$ then
            Apply Shapiro-Wilk on $S^*$, $S^* = \{S x_i\}$;
        else
            Apply Royston’s test on $S^*$, $S^* = \{S x_i\}$;
        end
    end
    Find $x_h$, such that $S^*$ has the best normality;
    Remove $x_h$ from $X$: $S\{S^* x_h\}$;
    $k: k-1$;
end

due to diet, physical and psychological changes, and other factors deviate the signal from Gaussianity. From this Gaussian model, we create the synthesized ECG heartbeats.

5.2.2 Parallel Classifiers to Reduce Instability

The main purpose of data generation is to increase biometric system performance by making use of the abundance of imposter dataset. The number of real genuine observations is small, we restricted it to 20 observations. On the other hand, we have thousands of imposter data. Due to small number of real genuine observations, classifiers’ structures change significantly depending on the imposter data that train the classifiers. We
propose a parallel classifier structure, and Figure 5.2 presents the scheme for it. All classifiers within the parallel classifier are trained with the same set of genuine training dataset but each classifier is trained with different sets of imposter data. The mean value of confidences of the classifiers’ outputs is used to make a classification decision.

5.3 Experimentation

This section investigates three main experiments: first, it presents biometric system improvement as a result of data synthesis, second experiment compares biometric system accuracy with data synthesis versus systems with different dimensionality reduction
techniques from the literature, and lastly, the third experiment demonstrates the parallel classifier performance. Throughout all experiments, the bagging classifier in Section 3.2.3 was used.

5.3.1 Synthesized ECG Heartbeat Generation

This experiment reports the improvement achieved in biometric system’s EER, TAR, and TRR quantities applying different parameters. Synthetic data were generated as explained in Section 5.2.1. The generated data were added to the pool of real genuine data, and they were used to train a bagging classifier. Table 5.2 presents an experiment when the real genuine data were restricted to the first 20 observations of every subject of the 1,012 subjects from UofTDB.

From Table 5.2, it can be noticed that the best EER from the examined parameters was achieved when we trained a classifier with 220 genuine observations (200 synthesized genuine data + 20 real genuine data) and 200 imposter data. Hence, this proves that adding data synthesis improves results. One may inquire why do we not consider the TAR of 400 synthesis data and 20 imposter data as the best result? The reason is TAR, unlike EER which considers both TAR and TRR, ignores TRR. TRR for the same experiment (400 synthesized data and 20 imposter data) has a significant drop from the average TRR of all experiments; it has a TRR of 82.28%. The reason behind that is that the classifier was biased towards the genuine data. It is worth mentioning that the reported TAR and TRR were calculated when the operating threshold that splits genuine from imposter classes for the bagging classifier was assigned to 50%.

We considered 20 real genuine observations as the baseline, but we also conducted other experiments when real genuine was 30 and 60 observations. Table 5.3 tabulates the EER that was achieved along with its corresponding number of synthesized data and number of imposter data. This table further confirms our hypothesis on the fact that adding generated synthesized data reduces data imbalance and constructs a better
Table 5.2: Experiment for 20 real genuine data. Synthetic genuine data were appended with real genuine data in training the classifier
classifier.

5.3.2 Comparison to Dimensionality Reduction

Dimensionality reduction is one of the most used techniques in the literature to deal with the small sample size problem [99, 100]. In this experiment, we compared the biometric system with data synthesis versus biometric systems with PCA, Probabilistic PCA [142], Isomap [143], Laplacian [144], and Local Linear Embedding (LLE) [145]. In all of these biometric systems, a real genuine data of 20 observations were used, and a wide range of numbers of imposter data and numbers of reduced dimensions were experimented. Table 5.4 tabulates the parameters that achieved the lowest EER while Figure 5.3 computes ROC curves for the biometric systems based on data synthesis and all dimensionality reduction techniques with parameters that achieved the lowest EER. It is pertinent to mention that all biometric systems were implemented in identical environment using same sets of real genuine and imposter observations.

5.3.3 Parallel Classifier

Bagging classifier has been investigated for ECG heartbeats due to its capability to reduce instability in predictors. Some instability still exist despite its reduction. This instability can especially be noticed on performance of individual subjects rather than considering hundreds of subjects when calculating biometric system performance using confusion matrix. Our proposed parallel classifier further reduces such instability by using a bagging classifier in a parallel scheme. Table 5.5 reports instability results and presents the influence of the parallel classifier on stabilizing it. It can be observed from Table 5.5 that when there is no parallel classifier, TAR would have a standard deviation of 6.52% per subject and TRR would have a standard deviation of 0.61% per subject while TAR would have a standard deviation of 1.94% per subject and TRR would have a standard deviation of 0.10% when parallel classifier was used. The only difference
Table 5.3: Experiment for 30 (top) and 60 (bottom) real genuine observations to: emphasize on performance improvement when we have 30 and 60 instead of 20 genuine (Table 5.2), show that data synthesize improves the results, and lastly, to observe the upward improvement when training sample size was increased.
<table>
<thead>
<tr>
<th>Method</th>
<th>EER (%)</th>
<th># of impos.</th>
<th># of dim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>9.92</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Prob. PCA</td>
<td>13.47</td>
<td>120</td>
<td>20</td>
</tr>
<tr>
<td>Isomap</td>
<td>16.16</td>
<td>120</td>
<td>10</td>
</tr>
<tr>
<td>LLE</td>
<td>14.82</td>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td>Laplacian</td>
<td>13.43</td>
<td>250</td>
<td>20</td>
</tr>
<tr>
<td>Data synthesis</td>
<td><strong>6.71</strong></td>
<td>200</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 5.4: Dimensionality reduction techniques best performances based on achieving the lowest EER among classifiers in the parallel classifier is that the imposter datasets is different in each classifier. Complexity can be an issue. If training a classifier takes $t$ seconds, then training $n$ parallel classifiers needs $n \times t$ seconds.

### 5.4 Summary

Two contributions have been proposed in this chapter: after analyzing Gaussianity of ECG observations, a proper and simple technique to generate ECG heartbeats synthesis was proposed, also an innovative methodology to reduce classifiers instability was presented. We used Sequential Forward Selection along with Shapiro-Wilk univariate and Royston’s multivariate normality tests to find a subset of ECG heartbeats variables that exhibit multivariate normal distribution. Our analysis suggests that more than 20 variables in the ECG heartbeats have multivariate normal distribution. Those multivariate variables capture the main features of ECG heartbeats. Therefore, they assist us in capturing the underlying Gaussianity of heartbeats and further support our hypothesis that ECG heartbeats exhibit a multivariate Gaussian distribution should deviating
Figure 5.3: ROC curves for the biometric system with different dimensionality reduction techniques and comparison to this chapter’s data synthesis approach. Dimensionality reduction techniques with lowest EER is plotted.

Factors did not occur. ECG heartbeats synthesis was used to generate genuine subject data to increase its sample size in a verification biometric system. When only 20 real genuine heartbeats were used and 200 synthesized heartbeats were generated, the biometric system achieved an EER of 6.71% in comparison to a minimum of 9.35% when data synthesis was not utilized. Biometric system with data synthesis outperformed several other biometric systems which employed dimensionality reduction techniques. Its EER outperformed PCA by 3.21%, Probabilistic PCA by 6.76%, Isomap by 9.45%, Local Linear Embedding by 8.11%, and Laplacian by 6.72%.
Table 5.5: Standard deviation of TAR and TRR for biometric systems with and without parallel classifier

Classifier instability is problematic especially when the sample size of the data is small. Bagging is usually used to reduce such effect, so we used it; however, to further reduce instability, we proposed a parallel classifier scheme. All classifiers were trained with the same set of genuine data and each classifier was trained with different set of imposter data. This method reduced classifier instability. Through this scheme, we could reduce TAR instability from 6.52% standard deviation to 1.94% standard deviation. The proposed contributions are expected to produce promising results in other applications.
Chapter 6

One-Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP)

Having access to only specific number of observations is a problem in several applications. Chapter 5 deals with such issue by synthesizing data. However, in some applications, adequate amount of data can be available. In fact, data may stream continuously in occasions such as wearable wrist and arm bands. We exploited the abundance of data to achieve a more robust biometric system, and for that we propose One-Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP) in this chapter and continuous authentication in the subsequent chapter.

1DMRLBP is inspired by computer vision widely examined Local Binary Patterns (LBP) which is a two-dimensional feature extraction. For ECG signal, we desired a feature type with the following properties:

- Fast to extract;
- Applicable to one-dimensional signal;
- Tolerable to various noise types that interfere with ECG signal;
- Provides discriminative feature extraction;
• Considers and extracts temporal variations among ECG heartbeats; and,

• Suitable for the proposed continuous authentication task.

This chapter briefly introduces LBP for images, expands on LBP features that considered extracting temporal variations, then explains 1DMRLBP. As explained in Section 1.4, there are several sources of noise that interfere with ECG signal collection; hence, we apply the preprocessing stage in Section 3.2.1 and the offline IQR outlier removal in Chapter 4, yet there still exist some scaling, quantization, and misalignment noise that affect biometric system performance. Also, most types of feature extraction techniques consider each heartbeat observation as an independent heartbeat without considering the possible temporal changes and variation that might occur on observations. For these issues, we studied and proposed 1DMRLBP.

6.1 Background of Local Binary Patterns

Before delving into the features we propose in this thesis, 1DMRLBP, we provide some background about the origin of these features, some advancements in the area, and trials to change it to be applicable to one-dimensional signals. One of the advantages is it considers extracting temporal variation among consecutive observations, and it extracts one feature vector from multiple observations. We also present literature review about LBP techniques that extract temporal information.

6.1.1 Review of Local Binary Patterns

In computer vision and image processing areas, LBP features were first introduced in [146], and they were originally proposed for texture classification. LBP has been examined in many image based applications. They are computationally inexpensive features to extract and has the capability to recognize corners, spots, edges, and straight
sections [146]. LBP also showed high tolerance to illumination changes [146]. For these reasons, LBP features have been used in object detection and recognition as the works including, but not limited to, [147, 148, 149, 150, 151].

In 2D images, the simple LBP feature extraction as explained in [146] works by isolating each \(3 \times 3\) pixels matrix, assumes the center pixel as the illumination pixel, and neglects it. The other eight pixels are considered as the texture of that specific patch. This texture information is encrypted and stored in the center pixel. The encryption occurs by subtracting the eight pixels from the center pixel, calculating \(\text{sign}(x)\), then changing the binary value to decimal. These steps are formulated in Equation (6.1). Based on the eight pixel encryption, \(2^8\) possible LBP values can be obtained from \(3 \times 3\) matrices.

\[
LBP(x_o, y_o) = \sum_{i=0}^{7} \text{sign}(g_i - g_o)2^i
\]  

\text{(6.1)}

Where \(LBP(x_o, y_o)\) is the LBP value for the center pixel in the \(3 \times 3\) matrix. The decimal value of \(LBP(x_o, y_o)\) represents the texture for this \(3 \times 3\) pixel window. \(g_i\) is the gray scale value of the surrounding pixels, and \(g_o\) is the gray scale value of the center pixel.

In Equation (6.1), \(\text{sign}(x)\) function was defined as:

\[
\text{sign}(x) = \begin{cases} 
1 & \text{if } x \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]  

\text{(6.2)}

An example of the procedure of the simple LBP feature extraction is illustrated in Figure 6.1.

There are many variations to LBP for 2D images. Circular LBP which considered a circular extraction rather than the square one was proposed in [152]. Multi-Block Local Binary Patterns (MBLBP) [147] extracted pixels’ LBP values then applied Haar-like wavelets on the pixels’ LBP values to extract features. Multiple single LBP features joined
Chapter 6. 1DMRLBP

Figure 6.1: Simple LBP feature. After subtracting the surrounding pixels from the centre pixel and taking their $\text{sign}(x)$ value, a stream of binary bits are concatenated from an arbitrary but consistent position.

Together using Sequential Forward Selection algorithm was proposed in [148]. Instead of subtracting pixels values from the center pixel, Center-Symmetric in [153] proposed to subtract each pixel from its opposite pixel. Rather than applying the LBP operation on raw pixel, Sobel-LBP [154] applied Sobel operation on the raw pixels then applied LBP on the Sobel extracted pixels. Furthermore, some works was conducted on the $\text{sign}(x)$ formula in Equation (6.2); the work in [155] proposed fuzzy operation instead of $\text{sign}(x)$ formula. The work in [156] proposed Local Ternary Patterns. Local Ternary Patterns were similar to LBP. It just modified the $\text{sign}(x)$ formula in Equation (6.2) such that if the surrounding pixel is equal, greater, or less than the value of the center pixel, it will be assigned with 0, 1, or -1, respectively. The work in [157] reviewed several other LBP advancement.

Recently, very few attempts were considered to apply LBP for one-dimensional (1D) signal processing as in [7, 158, 159, 160]. Both works, [158, 159], applied it to a 1D signal. However, the work in [160] projected an image to a 1D space by taking the cumulative sum in the x and y directions; hence, the 2D image was changed to a 1D image. Afterwards, One-Dimensional LBP (1DLBP) was applied on the projected 1D signal. The authors suggested that such projection has better discriminative capability than 2D LBP on images.

The work in [7] extracted the value of sample $x_c$ as:
\[ LBP(x_c) = \sum_{i=0}^{p-1} \text{sign}(x_{c+i} - x_c)2^i + \text{sign}(x_{c+i+1} - x_c)2^{i+p} \]  
(6.3)

\[ LBP(x_c) = \sum_{i=0}^{p-1} \text{sign}(x_{c+i-p-d+1} - x_c)2^i + \text{sign}(x_{c+i+d} - x_c)2^{i+p} \]  
(6.4)

Where \( p \) is the number of points (time-samples) to be considered on each side of \( x_c \), and \( x_c \) is the center time-sample that 1DLBP is desired to be extracted for. \( \text{sign}(x) \) is same as in Equation (6.2).

An illustration for the 1DLBP examined in [7] is presented in Figure 6.2.

\[ X(t) \]

![Diagram](image)

\[ \text{LSB} \quad \text{Thresholding} \]

\[ LBP = (10100111)_2 = (167)_{10} \]

Figure 6.2: 1DLBP used in [7]

### 6.2 Review of Temporal and Spatial Local Binary Patterns

Several 2D LBP variants attempted extracting temporal and spatial variations from observations. While they achieved promising results in their designated applications, they
Chapter 6. 1DMRLBP

were not suitable for the application proposed in this thesis. Being an offline feature extraction, has long feature vector, extracts features from predefined number of observations are among some of the reasons that make literature available LBP variants that consider temporal variations unsuitable for our application. Features are offline in a sense that they require future time observations in order to extract features at the current time.

In [161], the authors proposed Volume Local Binary Patterns (VLBP). VLBP feature type extracted LBP features for pixels from three frames, and it gave different weight to each of the three frames. From these frames, VLBP extracted features from co-occurred pixels on three orthogonal planes only. The resultant histogram would be of a length of $3 \times 2^p$ bins, where $p$ was the number of points to consider around a desired pixel. VLBP was proposed for offline applications (i.e. future observations should be available beforehand).

In [162], the authors acknowledged the long feature vector of [161] and proposed a fix to that issue. They introduced two techniques to reduce the number of bins, yet the issues of being an offline approach and an approach that considered only three frames persisted. First technique was to use six unique points from the three orthogonal planes. The second method dealt with reducing the number of features by taking the average value of orthogonal planes points.

The work in [163] proposed Spatio-Temporal Local Binary Patterns (STLBP). This type of features was online and applicable for real-time applications. For each pixel, it extracted the binary patterns for that observation using its current neighbors, then it extracted binary patterns from the same pixel location but from different observations. The feature vector was compiled of two histograms: one from the current observation and the other was from the previous observation.

The work in [164] extracted spatial and temporal LBP by prediction. It used previous frames to predict the value of a pixel in the current frame. It compared the predicted value against the actual value and incorporated that value in the LBP feature extraction
6.3 One-Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP)

1DMRLBP are inspired by the image based LBP [146] yet they are modified and enhanced to be applicable to 1D signals, to tolerate noise, to preserve ECG heartbeats morphology, and to account for ECG signal temporal variations through an extraction mechanism. Furthermore, 1DMRLBP features type is an online feature extraction in a sense that it can be applied in real-time, and it depends on past and current observations only.

1DMRLBP preserves texture of multiple time domain signals in one feature vector. Like 1DLBP and LBP, 1DMRLBP captures and encodes the shape of an ECG signal using binary patterns. Unlike 1DLBP, LBP, and many of their variations in the literature, 1DMRLBP extracts a feature vector that captures temporal changes of observations, and it has a mechanism to merge $n$ time domain heartbeats into 1 feature vector such that:

\[
1DMRLBP(x(t)) : \{x(t) \in \mathbb{R}^{k \times n}, 1DMRLBP(.) \in \mathbb{Z}^{m}\} \tag{6.5}
\]

where $n$ is number of heartbeat samples, $k$ is the dimensionality of time domain heartbeat sample, and $m$ is the dimensionality of the feature vector. Subsequent sections, in particular Algorithm 2, explains the transformation mechanism of Equation (6.5) that changes $n$ observations into one feature vector. Extracting 1DMRLBP is a low computational complexity operation; hence, it can be implemented in low power devices. The features and properties of 1DMRLBP are:

1DMRLBP Accounts for Quantization Error

In 2D images, pixel values have a known range (e.g. $[0,\ldots,255]$ in the case of gray scale image); hence, Equation (6.2) performs well when used. In such case, the influence of
step size is known (i.e. the difference between two consecutive levels is numerical 1 with a maximum difference of 255). In ECG signal, however, that’s not the case. Despite quantization level for the acquisition instrument could be known, the highest peak of the ECG signal stays unknown. Consequently, Equation (6.2) with one step size difference might not be adequate to capture the morphology of an ECG signal. Equation (6.2) is modified to account for such difference by adding a leeway in binary patterns extraction to accommodate such issue which cannot be overcome with the usual binary patterns extraction, which is based on hard decision. A parameter, $\epsilon$, is included in Equation (6.6) to account for quantization error. Also, adding this leeway reduces the influence of ripples (noise) on ECG signals.

$$\text{sign}(x) = \begin{cases} 
1 & \text{if } x + \epsilon \geq 0 \\
0 & \text{otherwise}
\end{cases} \quad (6.6)$$

1DMRLBP Captures Multiple Resolutions

Unlike 1DLBP in [7, 160], 1DMRLBP does not capture texture based on fixed number of points only. It considers different distances, $d$, and points, $p$, where $d$ is how far from the desired time-sample, $x(t)$, the features start to get extracted, and $p$ is how many time-samples are considered for 1DMRLBP feature extraction from each side. Figure 6.3 demonstrates the multi-resolution concept for the proposed binary patterns, $BP$.

$$BP(x(t)) = \sum_{i=0}^{p-1} \text{sign}(x(t + i - p - d + 1) - x(t))2^i + \text{sign}(x(t + i + d) - x(t))2^{i+p} \quad (6.7)$$

Where $\text{sign}(.)$ is as in Equation (6.6). $BP(x(t))$ is assigned to a value of zero when its parameters require information that is out of boundaries. It becomes out of boundaries when $t+i+d > k$, where $k$ is the heartbeat length, and when $t+i < p+d$. From Figure 6.3, with different $p$ and $d$, multiple resolutions can be captured for the same sample. Such
structure retrieves more signal characteristics than single resolution. Temporal variations and avoiding subsequent fusion stages are desired for the system in this.

1DMRLBP Considers Temporal Changes

In conventional LBP, and 1DLBP in particular, a feature vector is extracted from every observation. However, that is not the case for 1DMRLBP. 1DMRLBP accounts for observations’ temporal deviation through an algorithm that extracts one feature vector. This feature vector represents several consecutive observations that may undergo temporal changes. The advantage of this mechanism is that it results in one feature vector from multiple time-domain observations, and this characteristic is crucial for the CA system in this thesis.

1DMRLBP is desired to capture both local and global features of a signal. Also, it is preferred to have tolerance towards sparse noise and segments misalignment. Hence, similar to other LBP variants, BP values distribution is calculated. However, unlike other LBP feature extractions, 1DMRLBP extracts one feature vector from multiple observations.
To extract 1DMRLBP feature vector from $n$ observations: first, BP is extracted for all $n$ observations. Afterwards, we preassign overlapping and/or non-overlapping windows ranges. The windows have two parameters, window size, $w$; and shift, $\Delta$. At last, we apply the feature extraction mechanism that we propose in Algorithm 2 to extract one feature vector from $n$ observations. A visual example of mechanism that extracts temporal progression and extracts local and global feature extraction using six windows is illustrated in Figure 6.4.

\textbf{Algorithm 2:} 1DMRLBP feature extraction

\textbf{Data:} $X \in \mathbb{Z}^{k \times n}$, binary patterns for $n$ observations

$\Delta$, windows shifting steps, $w$, windows sizes

\textbf{Result:} $x \in \mathbb{Z}^{m \times 1}$, 1DMRLBP

Start with first observation $i = 1$, first window $j = 1$, and an empty feature vector of the 1DMRLBP = [ ];

Calculate $l$, number of windows, $l = 1 + \lfloor \frac{k-w}{\Delta} \rfloor$ ;

\textbf{while} $j \leq l$ \textbf{do}

\hspace{1em} Initiate $S_j \leftarrow [ ]$ ;

\hspace{1em} \textbf{while} $i \leq n$ \textbf{do}

\hspace{2em} $S_j \leftarrow w_{i,j}$, the $j^{th}$ window in the $i^{th}$ observation;

\hspace{2em} \textbf{end}

\hspace{1em} 1DMRLBP$_j \leftarrow$ distribution of $(S_j)$ ;

\hspace{1em} Normalize 1DMRLBP$_j$ ;

\hspace{1em} \textbf{end}

\textbf{return} Concatenate $(1DMRLBP_j)^{l}_{j=1}$ for all windows

From the algorithm, it can be deduced that 1DMRLBP accounts for temporal changes of observations seamlessly. The temporal changes consideration is achieved by modeling the distribution of binary patterns for a specific window for several consecutive observa-
Chapter 6. 1DMRLBP

Binary Pattens

1DMRLBP

1 2 n

Histogram(Sj)

Normalized
histogram(Sj)

w1,j w2,j wn,j

Sj=[w1,j w2,j … wn,j]

Normalized
histogram(S1)

Normalized
histogram(S2)

Normalized
histogram(Sj)

1DMRLBP

Figure 6.4: 1DMRLBP feature extraction depicts the feature extraction mechanism

1DMRLBP

tions. The result is always one feature vector. Since the feature extraction is based on finding the binary patterns distribution of each window for several observations rather than finding the distribution from every single observation. Statistically this larger sample size leads to a more descriptive model, hence, higher accuracy in 1DMRLBP based biometric system. This is empirically proven in Section 6.4.4. Extracting features using different window sizes allows 1DMRLBP to capture both local and global features. Most, if not all, LBP variants calculate binary patterns distribution for each observation separately. This loses possible observations’ temporal changes and produces a less descriptive distribution than 1DMRLBP because it is calculated from a smaller number of samples (i.e. one observation).

Through the stages that lead to the extraction of 1DMRLBP features, the discriminative power of this type of features is constructed by two stages: first, BP encodes the
micro-texture of time-samples $x(t)$. Micro-textures can be thought of as a template. The micro-textures have $2^p$ patterns and capture edges, flat areas, or special patterns as in Figure 6.5. Second, frequency of these patterns from local and global windows within the same observation, and frequency of the patterns through multiple observations emphasize on specific patterns of a subject, thus, increase 1DMRLBP discriminative power.

<table>
<thead>
<tr>
<th>$x(t)$ Value</th>
<th>Micro-texture patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1023</td>
<td><img src="image" alt="Micro-texture 1023" /></td>
</tr>
<tr>
<td>0</td>
<td><img src="image" alt="Micro-texture 0" /></td>
</tr>
<tr>
<td>31</td>
<td><img src="image" alt="Micro-texture 31" /></td>
</tr>
<tr>
<td>992</td>
<td><img src="image" alt="Micro-texture 992" /></td>
</tr>
<tr>
<td>911</td>
<td><img src="image" alt="Micro-texture 911" /></td>
</tr>
<tr>
<td>513</td>
<td><img src="image" alt="Micro-texture 513" /></td>
</tr>
</tbody>
</table>

Figure 6.5: An example of some micro-texture patterns that 1DMRLBP features capture

1DMRLBP Achieves Robustness Towards Noise

Throughout the explained sections on the parts that comprise 1DMRLBP features, theoretically, these features can tolerate noise due to their method of extraction. Apart from quantization error, noise that causes shifting and scaling of the signal, signal segmentation misalignment, and other noise that may defect some heartbeats are other problems that 1DMRLBP can tolerate. Similar concept to [146], we capture micro-texture patterns, $M$, revolving around each time-sample, $x(t)$, and we do that by calculating the joint distribution of the micro-texture patterns, $m$, for all $p$ points in the 1DMRLBP.
Let $k = 1 - p - d$, then we have

$$M(x(t)) = m(x(t), x(t + k), x(t + 1 + k), \ldots, x(t + p - 1 + k))$$  \hspace{1cm} (6.8)

Shift invariance is achieved by subtracting the center time-sample, $x(t)$, from neighboring time-samples. So the distribution becomes:

$$M(x(t)) = m(x(t), x(t + k) - x(t), x(t + 1 + k) - x(t), \ldots, x(t + p - 1 + k) - x(t))$$  \hspace{1cm} (6.9)

We assume the term $x(t)$ as carrying just the level or the average value of the micro-texture of that time-sample and no micro-texture information. So it is independent of the terms $x(t + i + k)$, where $i \in 0, 1, \ldots, p - 1$. Therefore, we can factor it out [146]. As a result, the distribution becomes:

$$M(x(t)) = m(x(t))m(x(t + k) - x(t), \ldots, x(t + 1 + k) - x(t), x(t + p - 1 + k) - x(t))$$  \hspace{1cm} (6.10)

Next, $M(x(t))$ is approximated by omitting the independent term $m(x(t))$. The micro-texture of $x(t)$, $M(x(t))$ is:

$$M(x(t)) \approx m(x(t + k) - x(t), x(t + 1 + k) - x(t), \ldots, x(t + p - 1 + k) - x(t))$$  \hspace{1cm} (6.11)

Despite the fact that the distribution is approximated, but doing so achieves shift invariant features. To construct scaling invariant features, Equation (6.6) is applied to every term in Equation (6.11). With the explained steps thus far, 1DMRLBP features are invariant to noise that causes shifting and scaling, and it can tolerate quantization error.

The micro-texture, $M(x(t))$, has a binary value, which then is encoded into a decimal value. This is the binary pattern, BP, of a sample. Instead of classifying each BP by itself as in [148], we applied histogram on overlapping and/or non-overlapping windows.
Chapter 6. 1DMRLBP

Histogram does not only capture morphological information but it eliminates spatial information, hence, reduces misalignment that may arise due to noise interference or signal segmentation error.

Lastly, we capture the distribution of multiple observations. Such a step captures temporal variation among ECG signals and may reduce the effect of abnormal ECG heartbeats. Defective heartbeats, which may occur in the middle of normal heartbeats while being able to pass the outlier removal, might not sabotage the classifier decision as much as if feature extractions and classification are applied on every observation.

To summarize 1DMLRLBP contributions and properties:

- Unknown signal amplitude, quantization error, shifting, scaling, and misalignment issues are tolerated.
- To extract 1DMLRLBP, four parameters are needed $d, p, w$ and $\Delta$.
- Different combinations of $d$ and $p$ capture different feature-space resolutions.
- 1DMLRLBP has a mechanism to extract temporal changes from observations, and it captures local and global morphological features. These properties along with sequential sampling construct the CA system proposed in this thesis.

6.4 Experimentation

We experiment 1DMLRLBP in several settings to measure and prove its capability. In Section 6.4.1, its discriminative power to extract meaningful features for ECG signals have been investigated for medical and biometric applications. Depending on 1DMLRLBP parameters, the feature vector can be lengthy, i.e. thousands of features. A study and comparison was conducted to examine whether dimensionality reduction can improve the results. This experiment is presented in Section 6.4.2. In Section 6.4.3, we present an experiment that investigated 1DMLRLBP feature extraction without classifiers.
in order to examine its discriminative power without classifiers advantages. Algorithm 2 is an important component of 1DMRLBP; hence, its influence has been investigated in Section 6.4.4. Lastly, the experiments would be short if these features were not compared to state-of-the-art method, and Section 6.4.5 does the comparison.

6.4.1 1DMRLBP as a Feature Extraction

The capability of using 1DMRLBP for ECG analysis was examined on two databases for two different applications to show that 1DMRLBP is not restricted for a single application or a database. The applications are medical and biometrics applications. In the medical application, classification of healthy versus non-healthy ECG heartbeats were conducted on the PTB database explained in Section 3.1.1. For the biometrics application, 1DMRLBP was examined on UofTDB in Section 3.1.2 for a verification biometric system. A verification system was designed since it is a required stage in a continuous authentication (CA) system which needs to verify the identity of a person or to flag the person as an intruder.

Experiment on PTB Database for Medical Application

In this experiment, 1,000 healthy ECG heartbeats and 1,000 non-healthy ECG heartbeats randomly selected from the database were used to train classifiers. 1DMRLBP feature vectors were extracted as explained in Algorithm 2. The number of observations in the algorithm was set to \( n = 1 \) because neither CA nor observation temporal consideration were desired to be examined in this experiment. The classifiers explained in Section 3.2.3 were deployed. By setting the operating threshold of choosing between classes (healthy and non-healthy) to 50\%, the results are tabulated in Table 6.1.

It can be observed that \( p \) was set to 4 in most cases in Table 6.1. The parameter \( p \) controls the number of histogram bins for each window. Number of histogram bins is \( 2^{2 \times p} = 256 \) bins for each window. The reason to slightly lower results when \( p = 5 \) is
Table 6.1: 1DMRLBP capability to classify ECG heartbeats. Multiple resolutions are read respectively, for example $p = (4, 4)$ and $d = (100, 200)$ means $p = 4$ and $d = 100$, then again $p = 4$ and $d = 200$. **Classifier**, **Points, $p$**, **Distance, $d$**, **Shift, $\Delta$**, **Win. size, $w$**, **Feat. vector length**, **FRR**, **FAR**, **PR**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Points, $p$</th>
<th>Distance, $d$</th>
<th>Shift, $\Delta$</th>
<th>Win. size, $w$</th>
<th>Feat. vector length</th>
<th>FRR</th>
<th>FAR</th>
<th>PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>256</td>
<td>0.52</td>
<td>0.06</td>
<td>0.83</td>
</tr>
<tr>
<td>Bagging (no $\epsilon$)</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1000</td>
<td>256</td>
<td>0.64</td>
<td>0.08</td>
<td>0.79</td>
</tr>
<tr>
<td>Bagging ($\epsilon = 0.01$)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1000</td>
<td>256</td>
<td>0.29</td>
<td>0.34</td>
<td>0.68</td>
</tr>
<tr>
<td>Bagging ($\epsilon = 0.02$)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1000</td>
<td>256</td>
<td>0.17</td>
<td>0.23</td>
<td>0.75</td>
</tr>
<tr>
<td>Bagging ($\epsilon = 0.05$)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1000</td>
<td>256</td>
<td>0.27</td>
<td>0.36</td>
<td>0.71</td>
</tr>
<tr>
<td>Bagging (no preproc.)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1000</td>
<td>256</td>
<td>0.17</td>
<td>0.23</td>
<td>0.75</td>
</tr>
<tr>
<td>Bagging ($\epsilon = 0.10$)</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1000</td>
<td>1,024</td>
<td>0.27</td>
<td>0.36</td>
<td>0.71</td>
</tr>
<tr>
<td>Bagging ($\epsilon = 0.20$)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1000</td>
<td>1,024</td>
<td>0.27</td>
<td>0.36</td>
<td>0.71</td>
</tr>
<tr>
<td>Bagging ($\epsilon = 0.30$)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1000</td>
<td>1,024</td>
<td>0.27</td>
<td>0.36</td>
<td>0.71</td>
</tr>
<tr>
<td>Bagging ($\epsilon = 0.40$)</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1000</td>
<td>1,024</td>
<td>0.27</td>
<td>0.36</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Note: $p = (4, 4)$ and $d = (100, 200)$ means $p = 4$ and $d = 100$, then again $p = 4$ and $d = 200$. **Classifier**
associated with the need for more training data than that when \( p = 4 \) since \( p = 5 \) has a feature vector of 10,240 features in comparison to 2,560 features when \( p = 4 \). However, when we examined a smaller \( p \) value, \( p = 2 \), the result did not reduce dramatically, but the feature vector was significantly reduced to 160 features. We used two sets of parameters throughout this work: \( d = (100, 200), p = (4, 4), \Delta = (100, 200), w = (500, 400) \), and \( d = (100, 200), p = (2, 2), \Delta = (100, 200), w = (500, 400) \). These parameters are relative to the ECG heartbeat length, \( k \). They can be presented as a proportion to \( k \) so they can be applicable to other signals with different sampling rate. For example, PTB heartbeats have 1,000 samples per heartbeat because the heartbeat is sampled at 1KHz, and we considered each heartbeat to span one second as explained in Section 3.2.1. When \( d = (100, 200), p = (4, 4), \Delta = (100, 200), w = (500, 400) \), and \( d = (100, 200), p = (2, 2), \Delta = (100, 200), w = (500, 400) \), the parameters can be expressed as \( d = (0.1 \times k, 0.2 \times k), p = (4, 4), \Delta = (0.1 \times k, 0.2 \times k), w = (0.50 \times k, 0.4 \times k) \), and \( d = (0.1 \times k, 0.2 \times k), p = (2, 2), \Delta = (0.1 \times k, 0.2 \times k), w = (0.50 \times k, 0.4 \times k) \), respectively. As can be seen from Table 6.1, these parameters achieved the highest performance. The plotted feature extraction for an ECG segment is shown in Figure 6.6. This plot shows the sparseness of the features. Around 90% of the features have a value of zero.

The following explains how the length of 2,560 was constructed. In case of \( d = 100, p = 4, \Delta = 100, w = 500 \), we have 6 windows calculated using \( l = 1 + \lfloor \frac{k - w}{\Delta} \rfloor \). Each window has \( 2^{2 \times p} = 256 \) bins. Windows’ feature vectors are concatenated to have \( 6 \times 256 = 1,536 \) features. Furthermore, in the next resolution when \( d = 200, p = 4, \Delta = 200, w = 400 \), there are 4 windows with 256 features for each window. All windows are concatenated to achieve \( 4 \times 256 = 1,024 \) features. The total is 2,560 features per feature vector.

It can be noticed from Table 6.1 that SVM has significantly worse performance than bagging. SVM with Gaussian Radial Basis Function (RBF) was used. Data over-fitting and bias are some of the reasons that achieved the low performance since training error
Figure 6.6: Feature extraction with parameters $d = (100, 200)$, $p = (4, 4)$, $\Delta = (100, 200)$, $w = (500, 400)$ for an ECG heartbeat.

for SVM was 0% while testing error was significantly higher. Signal instability is another reason for the bad performance of SVM. Bagging performance is in Table 6.1. The improvement of bagging over single decision tree was also experimented. Its result is reported in Table 6.1 clearly shows that bagging achieves better results than SVM and single decision tree.

The effect of modification by adding the leeway parameter $\epsilon$ in Equation (6.6) and the improvement due to the preprocessing stage are noted in Table 6.1. When choosing the value of $\epsilon$, we should consider that the ECG heartbeat might have some noise, and we do not want the noise to affect the BP extraction. This is an advantage over the hard binary patterns extraction in most LBP feature variants that are based on Equation (6.2). Also, $\epsilon$ value should not be of a high value that it can conceal intrinsic heartbeats features. We chose $\epsilon$ of a value proportional to the standard deviation of the examined signals; e.g. 10% of ECG signals standard deviation was $\approx 0.001$. Table 6.1 reports experiments with $\epsilon$ values of 0.001, 0.01, 0.02, and 0.05. It is worth noting that all other experiments
in this chapter were conducted with $\epsilon = 0.001$.

$ROC$ curve for $d = (100, 200), p = (4, 4), \Delta = (100, 200), w = (500, 400)$ is presented in Figure 6.7, which represents the performance of the system and illustrates that the curve does not undergo drastic changes when the operating point is changed.

Figure 6.7: $ROC$ for parameters $d = (100, 200), p = (4, 4), \Delta = (100, 200), w = (500, 400)$.

**Experiment on UofTDB for Biometrics Application**

After examining 1DMRLBP’s capability to extract ECG discriminative feature for medical application, it was experimented for a biometrics application. First part of this experiment examined 1DMRLBP parameters choice. It was experimentally discovered that the best results among the examined parameters were achieved using the same parameters that resulted in the highest performances in Section 6.4.1. An experiment examining the influence of the parameters was conducted again on UofTDB but now for a verification biometric system in order to confirm the parameters choice and to show
that having different resolutions affect the results. Result is plotted in Figure 6.8. In Figure 6.8, the highest performing two sets of parameters achieved almost identical accuracy; however, the set of parameters with $p = (2, 2)$ produced a feature vector of dimensions that is only 6.25% of that with the parameters using $p = (4, 4)$.

To prove the capability of 1DMRLBP to perform well for biometrics purposes, two types of experiments were designed for the task: one experiment used one session subjects in order to utilize the whole database of 1,012 subjects, hence, ensure scalability in low variance in performance and compare to state-of-the-art method that examined single session in the literature. The other experiment considered subjects with multiple sessions (82 subjects) to illustrate 1DMRLBP’s capability to be applied in real-life situations and to avoid possible classifier bias. The training and testing datasets are non-overlapping and their construction is discussed in Section 3.2.5. Figure 6.8 reports the ROC curves for 1,012 single session subjects biometrics experiment. Multiple sessions experiment was conducted, and its results was compared to single session experiment. Figure 6.9 illustrates the result. After presenting that 1DMRLBP performs well on multiple sessions database, the experiments onward will be diverted back to single session to obtain a robust results based on 1,012 subjects.

The capability of 1DMRLBP to extract high discriminative features and the ability to distinguish between two types of ECG heartbeats are apparent from Table 6.1. Apart from medical application, this same model can be used as quality measure to avoid passing non-healthy (irregularly shaped) ECG heartbeats to biometric system. Hence, this system is possible to be looked at as another outlier removal stage. If we are to look at this system as a quality measure system and to emphasize on its performance, all the examined segments healthy and non-healthy (regularly and irregularly shaped, respectively) were successfully extracted by QRS detector in Section 3.2.2.

To summarize the outcomes of this experiment:

- 1DMRLBP is capable of extracting ECG heartbeat discriminative features.
Chapter 6. 1DMRLBP

Figure 6.8: 1DMRLBP in a verification biometric system

- Multi-resolution increased performance accuracy by 16%.

- Ensemble classifier with decision tree achieved higher accuracy than SVM and single decision tree.

- Preprocessing increased accuracy by 7% \( FRR \), 5% \( FAR \), and 2% \( PR \) in comparison to the experiment without preprocessing that had \( FRR = 24\% \), \( FAR = 28\% \), \( PR = 73\% \).

- The modified \( \text{sign}(.) \) with \( \epsilon = 0.001 \) in Equation (6.6) improved the result by 12% \( FRR \), 11% \( FAR \) and 7% \( PR \) in comparison to the experiment that did not use \( \epsilon \) in the \( \text{sign}(.) \) which had \( FRR = 29\% \), \( FAR = 34\% \), \( PR = 68\% \).

- 1DMRLBP can be applied for biometrics applications. It achieved an EER of 7.89% on 1,012 single session database and an \( EER \) of 10.10% on 82 multiple sessions.
Figure 6.9: Multiple sessions versus single session experiment. EER for single session is 7.89% and for multiple sessions is 10.10%.

6.4.2 1DMRLBP with Different Dimensionality Reductions Techniques

Neighbors samples in time domain signals are highly correlated, so dimensionality reductions are usually utilized to reduce redundancy. Same concept applies to several binary patterns variants such as the works in [165, 166], and many other publications that apply dimensionality reduction to LBP features. This experiment has two merits: to empirically show that dimensionality reduction is not needed in 1DMRLBP, and to demonstrate the discriminative power of 1DMRLBP when compared to raw time-samples signals.

We investigated several dimensionality reductions techniques such as: PCA, Isomap [143], Laplacian [144], and LLE [145]. For all these reduction methods, the feature vector of
1DMRLBP (length 2,560 [167]) was reduced to 120 dimension feature vector. PCA with higher number of components, 500, was also experimented. Figure 6.10 shows the ROC curves comparison among all these techniques. Table 6.2 tabulates the EER results. During training, the feature extraction for each observation was extracted with \( n = 1 \) in Algorithm 2. Also, during testing mode, observations temporal information was extracted (i.e. \( n > 1 \) in Algorithm 2).

![Figure 6.10: Full 1DMRLBP feature vector versus dimensionally reduced 1DMRLBP](image)

From Table 6.2, 1DMRLBP without dimensionality reduction outperformed all other systems that attempted to reduce dimensionality. From this we conclude, bagging classification outperformed the advantage of dimensionality reduction.

From this experiment, it can be observed that:
### Table 6.2: EER (%) in different settings

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Reduction</th>
<th>Classifier</th>
<th>EER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1DMRLBP</td>
<td>None</td>
<td>Bagging</td>
<td>7.92</td>
</tr>
<tr>
<td>1DMRLBP</td>
<td>Isomap (120 feat.)</td>
<td>Bagging</td>
<td>13.81</td>
</tr>
<tr>
<td>1DMRLBP</td>
<td>Laplacian (120 feat.)</td>
<td>Bagging</td>
<td>11.13</td>
</tr>
<tr>
<td>1DMRLBP</td>
<td>LLE (120 feat.)</td>
<td>Bagging</td>
<td>11.30</td>
</tr>
<tr>
<td>1DMRLBP</td>
<td>PCA (120 feat.)</td>
<td>Bagging</td>
<td>10.12</td>
</tr>
<tr>
<td>1DMRLBP</td>
<td>PCA (500 feat.)</td>
<td>Bagging</td>
<td>11.47</td>
</tr>
</tbody>
</table>

- 1DMRLBP with bagging alone outperforms other dimensionally reduced feature vectors with PCA, Laplacian, Isomap, and LLE.

- 1DMRLBP represents ECG signal better than the raw samples.

#### 6.4.3 1DMRLBP Features Contribution to Accuracy

We desired to investigate whether 1DMRLBP feature type has a substantial contribution in improving accuracy of a biometric system, and it is not other factors only such as the bagging classifier. Therefore, two biometric systems were compared. In both systems, bagging classification was not used, and Euclidean distance was utilized instead. One system implemented 1DMRLBP features, and the other system used same set of heartbeats but without 1DMRLBP, raw time-samples. The parameters in Section 6.4.1 that yield 2,560 dimensions feature vector were used in the biometric system that utilized 1DMRLBP features. On the other hand, the raw time-samples has feature vectors of 200 samples. PCA with 10 components was used in both systems. The advantage of 1DMRLBP is apparent from Table 6.3 with an improvement of $\approx 15\%$ in $EER$. 
<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Feature Reduction</th>
<th>Classification</th>
<th>EER(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1DMRLBP</td>
<td>PCA (10 feat.)</td>
<td>Euclidean</td>
<td>20.42</td>
</tr>
<tr>
<td>None</td>
<td>PCA (10 feat.)</td>
<td>Euclidean</td>
<td>35.87</td>
</tr>
</tbody>
</table>

Table 6.3: 1DMRLBP features contribution to accuracy

6.4.4 Capability and Advantage of Temporal Variation Extraction

The 1DMRLBP feature extraction capability to capture observations’ temporal changes using the mechanism explained in Algorithm 2 was compared to 1DMRLBP where temporal consideration was not exploited i.e. Algorithm 2 with $n = 1$.

When temporal changes were not considered, majority voting decision-making scheme was deployed. No extra steps were required when 1DMRLBP utilized temporal changes consideration algorithm since the algorithm results in one feature vector. Majority voting was applied after extracting 1DMRLBP for each heartbeat segment individually.

Figure 6.11 illustrates the comparison. Such experiment was conducted by fixing all other 1DMRLBP parameters. We conclude that capturing observations’ temporal changes improve biometric system accuracy.

6.4.5 Comparison to State-of-the-Art

We compared the implemented biometric verification system to a state-of-the-art biometric verification system AC/LDA [3]. AC/LDA biometric system does not require signal segmentation, and it uses autocorrelation as features. It applies LDA for dimensionality reduction, and it uses Euclidean distance for classification. Figure 6.12 demonstrates the performance of 1DMRLBP using Algorithm 2 versus AC/LDA method.

Training to testing proportion of 80% to 20%, respectively, was used for AC/LDA. In
AC/LDA, majority voting was used for the decision making while that was not needed for 1DMRLBP. For heartbeats outliers elimination, offline IQR elimination as explained in Section 4.3 was utilized in 1DMRLBP, while autocorrelation coefficient based outlier elimination was used in AC/LDA. IQR outlier elimination is not applicable to AC/LDA since it requires the extraction of individual heartbeats, but AC/LDA is not based on isolated heartbeats. \textit{EER} was calculated to be 7.89\% for 1DMRLBP while it was calculated to be 12.30\% for AC/LDA. From this experiment, we deduce that a biometric system based on 1DMRLBP is capable to outperform a state-of-the-art work.
6.5 Summary

This chapter presented the One-Dimensional Multi-Resolution Local Binary Patterns. It is an online feature extraction that has a mechanism to extract temporal changes from observations, and it captures local and global morphological features. Regardless of the number of observations examined, one feature vector with a constant feature length is extracted. The classifier that classifies this feature vector does not need to know how many observations are used to construct it. These properties are crucial for the construction of the CA system proposed in Chapter 7. Due to 1DMRLBP method of extraction, unknown signal amplitude, quantization error, shifting, scaling, and misalignment issues are tolerated. The extraction of 1DMRLBP requires four parameters $d, p, w$ and $\Delta$, and different combinations of these parameter lead to different feature-space resolution.
Wide range of experiments were conducted to illustrate the viability of the 1DMRLBP features. 1DMRLBP was examined on single session dataset to achieve low performance variance scalability and was tested on multiple sessions dataset to prove its applicability in practical scenarios. 1DMRLBP could outperform a state-of-the-art work by achieving an EER or 7.89% in comparison to an EER of 12.30% in AC/LDA.
Chapter 7

Continuous Authentication

Password is an intuitive approach to prevent unauthorized people from accessing a specific media. Usually after logging-in to a system with a password, the system would either be logged-in until the user logs-out, or the system logs-out the user after a ‘grace period’ of inactivity. Several gaps exist in such authentication process. Some scenarios illustrating the gap include: an intruder can use the system in the absence of the genuine user if the system is accessed within the session’s grace period, and an intruder with the user’s password would be able to access the system at any time. As a result of these issues, research in continuous authentication (hereafter CA) systems has emerged. CA is referred to the task of continuously authenticating users while flagging intruders who try to access the system. There are several applications where CA can be utilized as a safeguard, such as computers [168], aircraft cockpits [169], cellphones, machines operating, cars, and other applications.

In this chapter, we introduce a novel continuous authentication system that in its core uses 1DMRLBP and sequential sampling. The system dynamically and based on previous observations only decides on the number of observations needed to commit a decision and adjusts biometrics thresholds. Experiments and comparison to state-of-the-art are presented to confirm our findings.
7.1 Introduction

One stage towards achieving a robust CA system is by replacing the traditional authentication access methods (e.g. passwords and tokens) with biometrics. For a CA system, it is desired to have a biometric system that does not require the cooperation of the subject (passive biometric), cannot be spoofed, and available to all living human being.

To recap from Chapter 1 on reasons behind selecting ECG is: first ECG is recorded by attaching sensors to the body. This method of acquisition makes ECG biometrics superior over several other biometric systems where acquiring the biometric signal may distract the user. For example, in face recognition, the user may need to look at the camera; in fingerprint recognition, the user may need to swipe his/her finger; and in speaker recognition, the user needs to speak. Another feature for ECG signal is that the signal is quasi-periodic; hence, observations are continuously available. Also its wide usage led to the availability of inexpensive sensors, and price is crucial in manufacturing commercial applications.

CA systems decide whether a subject is a genuine or an intruder subject by examining a pool of a classifier outputs collected from several observations. Each classifier is fed by an observation, heartbeat in this thesis; or a representation of the observation, feature extraction; and it outputs a confidence. The class, genuine or intruder, on which an observation belongs to is decided based on the classifier confidence. In statistics, including machine learning, the term garbage-in garbage-out is well understood, undesired output is produced when an unrepresentative input is supplied. Therefore, having undiscriminating features that do not have small intraclass variation and large interclass variation among subjects jeopardize the CA system performance.

To construct a CA system, it would be crucial to have an online feature extraction, and the proposed 1DMRLBP is an online feature extraction. Also CA makes a decision from a pool of classifier confidences using fusion methods, whereas that would not be the case when 1DMRLBP is used since only one feature vector is extracted regardless of the
number of observations examined.

A conventional method to implement a CA system is that the designer preallocates a segment size that corresponds to the number of confidences the CA needs in order to make a decision. The designer also pre-sets decision thresholds to decide on the class of the subject from each segment of confidences. In this chapter, we propose a novel method that allows us not to preallocate segment size and decision thresholds. This contribution uses sequential sampling to achieve a CA system that utilizes 1DMRLBP features properties. Both 1DMRLBP and sequential sampling construct a CA system that adjusts the segment size and decision thresholds to make a decision while a biometric system keeps collecting input data.

7.2 Methodology

We use sequential sampling to achieve a system that adaptively changes segment size and decision threshold. The theory of sequential sampling is explained then experimentation is conducted to examine its applicability.

7.2.1 Sequential Sampling

Conventionally, CA systems use a predefined number of observations along with pre-set decisions threshold to make a positive (authenticate) decision or a negative (reject) decision. Setting the number of observations and decision threshold parameters has a trade-off between accuracy and time. Hence, a method that can avoid setting these parameters is desired. We examined sequential sampling [170] for that purpose. Sequential sampling has been mostly involved in quality control projects where the number of samples to be examined is unknown.

Sequential sampling is a statistical approach that follows criteria to accept a hypothesis, reject a hypothesis, or ask for more samples to decide. In a similar manner, we use
sequential sampling for ECG signal CA system such that the three criteria are: authenticate region, reject region, and continue region. Along with these criteria, we propose using ‘segments’. The onset of the first segment is on the first observation tested while the offset is when a decision is made. The second segment onsets on the first observation after the offset of the previous segment, and so on. The temporal information for observations are extracted using 1DMRLBP for the period of each segment only.

During the operation, when authenticate region criterion is met (positive authentication), it means a genuine person is providing his/her signal. Hence, CA system ends the current segment and initiates another one. On the other hand, when reject region criterion is met (negative authentication), it means an imposter or intruder has provided his/her signal. Therefore, CA ends the segment and flags the signal provider as an intruder. Lastly, when CA is in the continue region, it means a bigger sample size is needed. Figure 7.1 illustrates a hypothetical example of CA system in action using sequential sampling. It shows 10 segments where each segment has its criterion getting updated progressively as the segment’s sample size increases. The figure also shows the segment re-initialization once a criterion is met. The upper lines represent the minimum classification confidence to accept a sample while the lower lines represent the minimum classification confidence in order not to reject a sample. As the number of samples increases, the acceptance and rejection thresholds change accordingly (upper and lower lines in Figure 7.1, respectively). Unlike traditional works, an acceptance/rejection threshold is not set.

The main advantage of sequential sampling is that it is optimized to minimize the sample size needed to make a decision. Sequential sampling has two alternative hypotheses where each hypothesis is a one-sided hypothesis. When a sample is not significantly different from both hypotheses, a decision cannot be made and more observations are needed. Since a single confidence value is needed to check significance from the hypotheses, then the usefulness of the property of 1DMRLBP that achieves one feature vector
Figure 7.1: A hypothetical example illustrating how CA works. This CA model was trained with a subject and random observations were examined. Each triangle corresponds to an observation. The line that connects a set of triangles represents the segment that was used to reach a decision. The upper and the lower lines illustrate the decision threshold variations for that segment.

from multiple observations becomes crucial in this stage. One of the sequential sampling conditions is that the data should be drawn from a binomial distribution. The hypotheses for the sequential sampling are constructed from two distributions: genuine confidences and imposter confidences. Figure 7.2 illustrates the hypotheses.

The parameters that control the sequential sampling criteria are: Type I error, $\alpha$, Type II error, $\beta$; and the hypotheses means, $\mu_1, \mu_2$ for the imposter and genuine scores distributions, respectively. It may seem confusing that $\alpha$ and $\beta$ are set separately and unrelated to each other. This phenomena is due to the fact that we have two independent alternative hypotheses. On segment onset, the alternative hypotheses are set as in
Equation (7.1) [28]:

\( H_1 : \text{classification distance } \geq \mu_1 \)  
\( H_2 : \text{classification distance } \leq \mu_2 \)  

Setting up \( \alpha \) and \( \beta \) is setting up risks such that \( \alpha \) corresponds to the risk of rejecting \( H_1 \) and accepting \( H_2 \) while the sample belongs to \( H_1 \). Similarly, \( \beta \) corresponds to rejecting \( H_2 \) and accepting \( H_1 \) while the sample is actually from \( H_2 \).  

After setting up the hypotheses, the decision criteria needs to be updated based on the number of samples observed. The criteria are represented by the parallel lines separating the regions as seen in Figure 7.3. The equations for the lines are as in Equation (7.2).
Chapter 7. Continuous Authentication

Upper Line (UL) : \[ Y = bn + h_1 \] (7.2)
Lower Line (LL) : \[ Y = bn + h_2 \]

The slope, \( b \), can be calculated as \( b = \frac{\mu_1 + \mu_2}{2} \). The y-intercept can be calculated from the following two equations [28]:

\[
h_1 = \frac{B \sigma^2}{\mu_1 - \mu_2}, h_2 = \frac{A \sigma^2}{\mu_2 - \mu_1}
\] (7.3)

Where \( A = \log \frac{1-\alpha}{\beta} \), and \( B = \log \frac{1-\beta}{\alpha} \). \( \sigma \) is the standard deviation of the distances.

To compute the decision up to observation \( n \), \( c_n \), and check whether it meets any of the criteria then Equation (7.4) is used.

\[ C_n = \sum_{1}^{n} c(n) \] (7.4)

\( c(n) \) is the classification confidence for 1DMRLBP feature vector, which is calculated using Algorithm 2 for sample \( n \). Figure 7.3 illustrates a sequential sampling example in accepting a set of observations as genuine. When \( C_n \) was in the continue region, it requested more observations. It needed four observations to reach to the authenticate region.

It is worth mentioning that the wider the undecided area (the area between the lines in Figure 7.3) the more aggressive the criteria are. By aggressive we mean more observations are needed to make a decision. The closer the means are in the hypotheses, Figure 7.2, the more aggressive the criteria are. Same concept applies to \( \alpha \) and \( \beta \) such that the smaller the number the more aggressive the criteria are.

Sequential sampling criteria are more aggressive in a sense that the criteria are more difficult to meet. The more aggressive the criteria are the larger sample size that is required. Figure 7.4 illustrates the criteria with different parameters with fixed Type I and Type II errors while Figure 7.5 shows the criteria with different Type I and Type
II errors with fixed *means* values. For comparison purposes, line 4 in Figure 7.4 is the same as line 7 in Figure 7.5.

Referring to Figure 7.2, when we add a positive number to $\mu_1$ and subtract the same positive number from $\mu_2$, the distributions become closer to each other, and intuitively, the classes become harder to distinguish. Hence, the criteria is more aggressive and more observations are required to reach a decision. This is depicted in line 7 in Figure 7.4. On the other hand, when we subtract a positive number from $\mu_1$ and add a positive number to $\mu_2$, the *means* become further apart. For that, it would be easier to distinguish the classes, and accordingly, a criterion is easier to meet. This is illustrated in line 1 in Figure 7.4.
7.3 Experimentation

7.3.1 Comparison to State-of-the-Art CA Systems and Influence of Sequential Sampling Parameters

A CA system was implemented using sequential sampling as explained in Section 7.2.1. This system was examined for biometric application on UofTDB and was compared to other two continuous authentication systems of Ahmed et al. [1] and Labati et al. [2]. In this experiment, we also compared two 1DMRLBP based continuous authentication systems, one was based on 1DMRLBP that extracts temporal variations and sequential sampling and the other one was based on 1DMRLBP that does not deploy temporal variation [167](i.e. $n = 1$ in Algorithm 2) and sequential sampling. Lastly, we added sequential sampling to AC/LDA [3] state-of-the-art ECG biometric features. The purpose of the first experiment is to present performance improvement of the proposed CA system over state-of-the-art works. The purpose of the second experiment is of three folds: first is to show that a system with 1DMRLBP with temporal variation extraction and sequential sampling outperforms 1DMRLBP with no temporal variation extraction and sequential sampling. Also, to conclude that adding sequential sampling improves biometric results over the system with no sequential sampling as in Section 6.4.4; hence, sequential sampling purpose is not only for continuous authentication. The last purpose is to show that sequential sampling is applicable to 1DMRLBP that does not consider temporal variations. In implementing sequential sampling with AC/LDA, we desire to illustrate that sequential sampling application is not restricted to 1DMRLBP features and to reinforce that it adds biometric improvement when compared to the same system but without sequential sampling.

Different parameters set up change the criterion of the sequential sampling. The results in Table 7.1 present performance with different parameters. These results were collected by training a model for each subject. Afterwards, the model was examined on
all 1,012 database subjects, and FAR and FRR were collected. It can be noticed from Table 7.1 as how the FAR and FRR decrease as the sequential sampling criteria become more aggressive.

The choice of what parameters to use would be an application dependent. For an example, comparing tele-medicine application where ECG is monitored remotely and collected for several minutes to a computer access where user needs to log in as soon as possible. A heartbeat can be acquired at a rate of 1-1.5 heartbeat/second. If CA7 in Table 7.1 is considered, then around 30 observations are required to meet a criterion, which translates to around 24 seconds of ECG acquisition in comparison to 4 seconds in the case of CA1.

A state-of-the-art work of Ahmed et al. [1] was one of the systems we adapted for ECG biometrics and was compared to the proposed CA system. In [1], proportion instead of means were used in the construction of the sequential sampling criteria. The equations for the criteria that separate the three regions are [1]:

\[
UpperLine = -h_1 + sl \times n \\
LowerLine = h_2 + sl \times n
\]

where

\[
h_1 = \frac{1}{g} \left[ \log \left( \frac{1 - \alpha}{\beta} \right) \right], \quad h_2 = \frac{1}{g} \left[ \log \left( \frac{1 - \beta}{\alpha} \right) \right], \quad g = \log \left[ \frac{p_2(1-p_1)}{p_1(1-p_2)} \right], \quad and \quad sl = \frac{1}{g} \left[ \log \left( \frac{1-p_1}{1-p_2} \right) \right].
\]

\( p_1 \) corresponds to the maximum probability to flag an imposter, and \( p_2 \) denotes the minimum probability before authenticating a genuine user. No hypothesis generation was needed, and the criteria were generated using different \( p_1 \) values while setting \( p_2 = 1 - p_1 \) [1].

Table 7.2 presents the results. As it can be noticed from Table 7.2, changing \( p_1, \alpha, \) and \( \beta \) influence the results. Ahmed et al. [1] investigated these parameters in depth. However, the same concept of criteria aggressiveness explained in Section 7.2.1 regarding the width between the criteria lines applies to the work in [1].

By comparing Table 7.1 to Table 7.2, it can be noticed that on a fixed value of \( \alpha \)
<table>
<thead>
<tr>
<th>CA number</th>
<th>$H_1$ mean</th>
<th>$H_2$ mean</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>FRR</th>
<th>FAR</th>
<th>$IO$</th>
<th>$GO$</th>
<th>FRR</th>
<th>FAR</th>
<th>$IO$</th>
<th>$GO$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA1</td>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
<td>0.01</td>
<td>0.01</td>
<td>1.13</td>
<td>8.11</td>
<td>5.83</td>
<td>5.67</td>
<td>3.30</td>
<td>9.10</td>
<td>6.27</td>
<td>6.27</td>
</tr>
<tr>
<td>CA2</td>
<td>$\mu_1 - 0.1\sigma$</td>
<td>$\mu_2 + 0.1\sigma$</td>
<td>0.01</td>
<td>0.01</td>
<td>1.55</td>
<td>8.98</td>
<td>5.25</td>
<td>5.08</td>
<td>3.54</td>
<td>9.95</td>
<td>5.65</td>
<td>5.49</td>
</tr>
<tr>
<td>CA3</td>
<td>$\mu_1 - 0.3\sigma$</td>
<td>$\mu_2 + 0.3\sigma$</td>
<td>0.01</td>
<td>0.01</td>
<td>1.93</td>
<td>9.68</td>
<td>4.46</td>
<td>4.38</td>
<td>3.93</td>
<td>8.14</td>
<td>4.76</td>
<td>4.58</td>
</tr>
<tr>
<td>CA4</td>
<td>$\mu_1 + 0.1\sigma$</td>
<td>$\mu_2 - 0.1\sigma$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.85</td>
<td>7.51</td>
<td>6.51</td>
<td>6.32</td>
<td>3.93</td>
<td>8.14</td>
<td>4.76</td>
<td>4.58</td>
</tr>
<tr>
<td>CA5</td>
<td>$\mu_1 + 0.3\sigma$</td>
<td>$\mu_2 - 0.3\sigma$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.81</td>
<td>5.94</td>
<td>8.81</td>
<td>8.79</td>
<td>2.81</td>
<td>7.24</td>
<td>9.80</td>
<td>10.14</td>
</tr>
<tr>
<td>CA6</td>
<td>$\mu_1 + 0.4\sigma$</td>
<td>$\mu_2 - 0.4\sigma$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.68</td>
<td>4.82</td>
<td>11.01</td>
<td>11.19</td>
<td>1.59</td>
<td>5.71</td>
<td>12.18</td>
<td>13.07</td>
</tr>
<tr>
<td>CA7</td>
<td>$\mu_1 + 0.6\sigma$</td>
<td>$\mu_2 - 0.6\sigma$</td>
<td>0.01</td>
<td>0.01</td>
<td>0.39</td>
<td>1.57</td>
<td>30.10</td>
<td>34.38</td>
<td>0.92</td>
<td>3.17</td>
<td>34.50</td>
<td>40.40</td>
</tr>
<tr>
<td>CA8</td>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
<td>0.001</td>
<td>0.001</td>
<td>0.56</td>
<td>6.77</td>
<td>7.70</td>
<td>7.74</td>
<td>1.26</td>
<td>7.27</td>
<td>8.52</td>
<td>8.75</td>
</tr>
<tr>
<td>CA9</td>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.60</td>
<td>5.47</td>
<td>9.36</td>
<td>9.25</td>
<td>1.26</td>
<td>7.27</td>
<td>8.52</td>
<td>8.75</td>
</tr>
<tr>
<td>CA10</td>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
<td>0.01</td>
<td>0.001</td>
<td>0.32</td>
<td>9.33</td>
<td>7.71</td>
<td>5.80</td>
<td>1.41</td>
<td>12.03</td>
<td>5.61</td>
<td>3.56</td>
</tr>
<tr>
<td>CA11</td>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
<td>0.001</td>
<td>0.01</td>
<td>0.89</td>
<td>6.22</td>
<td>5.91</td>
<td>7.80</td>
<td>1.41</td>
<td>12.03</td>
<td>5.61</td>
<td>3.56</td>
</tr>
<tr>
<td>CA12</td>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
<td>0.1</td>
<td>0.01</td>
<td>1.41</td>
<td>12.03</td>
<td>5.61</td>
<td>3.56</td>
<td>1.41</td>
<td>12.03</td>
<td>5.61</td>
<td>3.56</td>
</tr>
<tr>
<td>CA13</td>
<td>$\mu_1$</td>
<td>$\mu_2$</td>
<td>0.01</td>
<td>0.10</td>
<td>3.30</td>
<td>7.386</td>
<td>3.71</td>
<td>5.40</td>
<td>3.30</td>
<td>7.386</td>
<td>3.71</td>
<td>5.40</td>
</tr>
</tbody>
</table>

Table 7.1: Sequential sampling performance with different parameters and comparison between sequential sampling with 1DM-RLBP features that considers temporal variation with 1DMRLBP that does not consider temporal variation. Figure 7.4 and Figure 7.5 visualize the criteria in this experiment.
Chapter 7. Continuous Authentication

<table>
<thead>
<tr>
<th>#</th>
<th>$p_1$</th>
<th>$\alpha = \beta$</th>
<th>FRR</th>
<th>FAR</th>
<th>IO</th>
<th>GO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.01</td>
<td>8.55%</td>
<td>9.30%</td>
<td>6.65</td>
<td>6.23</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.01</td>
<td>5.75%</td>
<td>8.82%</td>
<td>9.29</td>
<td>9.88</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.001</td>
<td>4.06%</td>
<td>7.33%</td>
<td>12.61</td>
<td>12.32</td>
</tr>
<tr>
<td>4</td>
<td>0.35</td>
<td>0.01</td>
<td><strong>4.33%</strong></td>
<td><strong>7.12%</strong></td>
<td><strong>14.22</strong></td>
<td><strong>14.19</strong></td>
</tr>
<tr>
<td>5</td>
<td>0.45</td>
<td>0.01</td>
<td>4.05%</td>
<td>10.33%</td>
<td>35.32</td>
<td>35.45</td>
</tr>
</tbody>
</table>

Table 7.2: Performance of Ahmed et al. [1] CA system

and $\beta$, the proposed CA consistently achieved lower FRR and FAR, and it could reach a decision with less number of observations. Furthermore on a close GO and IO values as in CA1 in Table 7.1 and #1 in Table 7.2, the proposed CA outperformed Ahmed et al. [1]. Also, the minimum FAR and FRR reported in the proposed method was not achievable in Ahmed et al. [1]. Hence, the proposed CA system performance surpasses a state-of-the-art from all examined aspects.

Another comparison to a state-of-the-art approach by Labati et al. [2] was conducted. The work in Labati et al. [2] applied different fusion techniques such as mean, median, 75th percentile, 90th percentile, 95th percentile, and max on a predefined number of observations. 1DMRMLBP with n=1 as feature extraction and bagging as a classifier were used in the experiment. The main purpose of this experiment was to compare sequential sampling to fusion in implementing a CA system. Table 7.3 presents the results for different segment sizes.

In the work of Labati et al., we do not have a dynamic number of GO and IO since a predetermined number of observations are used. If we take 5 observations for example, then our approach achieved FRR of 1.13% and FAR of 8.11% (Table 7.1) in comparison to an FRR of 3.93% and FAR of 9.69% (Table 7.3) to that using Labati et al. mean fusion method. Similarly, other comparisons can be performed, and we can conclude that using
Table 7.3: Performance (%) of 1DMRLBP with n=1 using continuous authentication methods in Labati et al. [2]

<table>
<thead>
<tr>
<th>Seg. size</th>
<th>Mean FAR</th>
<th>FRR</th>
<th>Median FAR</th>
<th>FRR</th>
<th>Max FAR</th>
<th>FRR</th>
<th>75th perc. FAR</th>
<th>FRR</th>
<th>90th perc. FAR</th>
<th>FRR</th>
<th>95th perc. FAR</th>
<th>FRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>9.69</td>
<td>3.93</td>
<td>10.84</td>
<td>5.95</td>
<td>34.12</td>
<td>0.50</td>
<td>21.19</td>
<td>1.11</td>
<td>34.12</td>
<td>0.50</td>
<td>34.12</td>
<td>0.50</td>
</tr>
<tr>
<td>10</td>
<td>8.71</td>
<td>2.36</td>
<td>9.3</td>
<td>2.56</td>
<td>44.63</td>
<td>0</td>
<td>21.39</td>
<td>0.20</td>
<td>36.32</td>
<td>0</td>
<td>44.63</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>8.15</td>
<td>2.06</td>
<td>9.12</td>
<td>2.61</td>
<td>50.11</td>
<td>0</td>
<td>19.74</td>
<td>0.55</td>
<td>35.94</td>
<td>0.13</td>
<td>45.25</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>7.81</td>
<td>0.77</td>
<td>8.40</td>
<td>1.35</td>
<td>54.78</td>
<td>0</td>
<td>19.94</td>
<td>0.19</td>
<td>35.86</td>
<td>0</td>
<td>47.24</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>7.38</td>
<td>1.00</td>
<td>7.98</td>
<td>1.50</td>
<td>60.53</td>
<td>0</td>
<td>20.04</td>
<td>0.25</td>
<td>35.59</td>
<td>0</td>
<td>47.36</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>7.08</td>
<td>0.35</td>
<td>7.66</td>
<td>0.70</td>
<td>63.14</td>
<td>0</td>
<td>218.91</td>
<td>0</td>
<td>34.71</td>
<td>0</td>
<td>46.10</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 7.4: Sequential sampling performance AC/LDA [3]

<table>
<thead>
<tr>
<th>CA number</th>
<th>Parameters</th>
<th>AC/LDA [3] with sequential sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA1</td>
<td>( \mu_1 ) ( \mu_2 )</td>
<td>0.01      0.01      5.88%    8.47%   18.11  17.13</td>
</tr>
<tr>
<td>CA2</td>
<td>( \mu_1 - 0.1\sigma ) ( \mu_2 + 0.1\sigma )</td>
<td>0.01      0.01      6.95%    4.98%   25.43  19.17</td>
</tr>
<tr>
<td>CA3</td>
<td>( \mu_1 + 0.1\sigma ) ( \mu_2 - 0.1\sigma )</td>
<td>0.01      0.01      6.59%    11.11%  14.17  14.61</td>
</tr>
<tr>
<td>CA4</td>
<td>( \mu_1 + 0.3\sigma ) ( \mu_2 - 0.3\sigma )</td>
<td>0.01      0.01      5.44%    13.68%  9.76   11.16</td>
</tr>
<tr>
<td>CA5</td>
<td>( \mu_1 ) ( \mu_2 )</td>
<td>0.001     0.001     6.67%    5.16%   24.79  19.58</td>
</tr>
<tr>
<td>CA6</td>
<td>( \mu_1 ) ( \mu_2 )</td>
<td>0.0001    0.0001   10.76%   5.16%   30.28  20.08</td>
</tr>
<tr>
<td>CA7</td>
<td>( \mu_1 ) ( \mu_2 )</td>
<td>0.01      0.001    9.46%    5.65%   24.78  17.14</td>
</tr>
<tr>
<td>CA8</td>
<td>( \mu_1 ) ( \mu_2 )</td>
<td>0.001     0.01     4.11%    11.80%  18.15  19.58</td>
</tr>
<tr>
<td>CA9</td>
<td>( \mu_1 ) ( \mu_2 )</td>
<td>0.1       0.01     9.65%    6.36%   17.91  11.36</td>
</tr>
<tr>
<td>CA10</td>
<td>( \mu_1 ) ( \mu_2 )</td>
<td>0.01      0.10     3.37%    21.76%  9.97   16.97</td>
</tr>
</tbody>
</table>
unset sample size with sequential sampling outperforms fusion method.

Table 7.1 reveals that 1DMRLBP that extracts temporal variation consistently outperforms 1DMRLBP without temporal variations extraction. Therefore, the improvement achieved in considering 1DMRLBP in Section 6.4.4 propagates even after adding a sequential sampling. We can conclude that considering temporal variation enhances performance. Furthermore, if we compare 1DMRLBP with sequential sampling and without sequential sampling in both cases: extracting temporal variations or not extracting temporal variations, in other words compare Table 7.1 to Figure 6.11, the improvement of biometric performance after adding sequential sampling is apparent on both types of features.

Lastly, sequential sampling is not applicable to 1DMRLBP features only. Table 7.4 presents the results when sequential sampling was applied to AC/LDA features. Adding sequential sampling also improved the performance of AC/LDA when compared to same features but without sequential sampling in Figure 6.12. On the other hand, comparing Table 7.1 versus Table 7.4, we can conclude that 1DMRLBP with sequential sampling outperforms AC/LDA with sequential sampling. For a fair comparison, one needs to compare the experiments where approximately same number of observations were used to make a decision. It can be noticed that 1DMRLBP with sequential sampling, whether we use \( n=1 \) or extract temporal variation, outperforms the performance of AC/LDA. One important point to note while comparing the results is that CA2 and CA3 in Table 7.1 are to be compared to CA3 and CA4 in Table 7.4, respectively. The reason behind this is that AC/LDA is based on distance measure while 1DMRLBP is based on confidence measure. The smaller the distance the higher the similarity while the higher the confidence the higher the similarity to a class. As result, when we add a positive number to \( \mu_1 \) (imposter \textit{mean}) and subtract a positive number from \( \mu_2 \) (genuine \textit{mean}), we are making the \textit{means} further apart in the AC/LDA experiment. Hence, the criteria become less aggressive. For this reason, CA4 in Table 7.4 require less number of \textit{IO} and \textit{GO} than CA1. This is
identical to adding a positive number to $\mu_2$ (genuine *mean*) in 1DMRLBP and subtracting a positive number from $\mu_1$ (imposter *mean*). Not all parameters in Table 7.1 could be applied to AC/LDA because doing so requires a number of observations that exceeds the number of the database’s subjects observations.

From this experiment, we conclude:

- Sequential sampling can be incorporated with 1DMRLBP and other features to achieve a CA system.
- Sequential sampling parameters control the aggressiveness of the CA system.
- Sequential sampling, apart from its advantage in achieving a CA system, it enhances biometric performance.
- The more aggressive the system is, the higher the accuracy yet the higher IO and GO.
- The mechanism of feature extraction in Algorithm 2 is essential to sequential sampling algorithm since it results in only one feature vector from multiple observations; hence, no fusion or decision-making algorithms are needed.
- The proposed CA system outperforms state-of-the-art continuous authentication systems.

### 7.3.2 Continuous Authentication in Action

This experiment mimics CA operation in a real-life scenario. In real-life, CA continuously examines segments. Every segment ends by making a decision. To simulate a scenario, we constructed a lengthy stream of observations from different subjects, and we trained a subject. The subject’s training and testing data did not overlap. The subject we trained was considered genuine while the others were considered imposters. Figure 7.6 simulates a CA system in action.
For this experiment, CA1 in Table 7.1 was used. When genuine observations were provided, authenticate criterion was reached while when imposters observations were encountered the reject criterion was met. Furthermore, in some segments, noise interfered and that required a bigger sample size to make a decision. This is obvious from segment 4 and segment 8 in Figure 7.6. Also, these two segments lie at the boundaries where imposter and genuine observations were mixed within one segment, which is a possible real-life scenario.

We can conclude from this experiment that this CA system can tolerate noise and can change sample size and decision thresholds during run-time based on previous observations confidences.

7.4 Summary

A continuous authentication system that uses sequential sampling is proposed. Sequential sampling brought advantages such that the system does not require a predefined number of observations in order to make a decision. The proposed CA keeps requesting observations until one of its two criteria is met. The two criteria are either to reject or to accept the examined segment of classifiers confidences. Also, sequential sampling does not need a pre-set decision threshold. The thresholds are adaptive and change in run-time. The CA system is set up by assigning theoretical Type I and Type II errors values which control the aggressiveness of the sequential sampling.

Aggressiveness is defined by the width of the separation between the maximum confidence before rejecting a hypothesis as being faulty and the minimum threshold to accept a hypothesis that the claimed identity is genuine. The wider the gap, the more aggressive sequential sampling is, which usually leads to the need of larger segment size which in turns achieves a higher accuracy. Therefore, the choice of parameters is an application dependent. There is a trade-off, the bigger the separation between the lines (i.e. more
aggressive), the higher the accuracy but more observations are needed (i.e. larger $IO$ and $GO$). Therefore, the choice of parameters is application dependent.

1DMRLBP features that extract temporal variations and extract single feature vector achieved the highest accuracy when compared to 1DMRLBP without temporal variation extraction. Also, CA system with sequential sampling was compared and outperformed state-of-the-art works. Lastly, CA real-life simulation experiment was presented. This simulation well explained the phenomena of dynamic segment size change and threshold in the presence of noise.
Figure 7.4: Sequential sampling criteria with different parameters. $\alpha = 0.01$ and $\beta = 0.01$. The parameters based on the line number are: (1) $\mu_1 + 0.4\sigma, \mu_2 - 0.4\sigma$, (2) $\mu_1 + 0.3\sigma, \mu_2 - 0.3\sigma$, (3) $\mu_1 + 0.1\sigma, \mu_2 - 0.1\sigma$, (4) $\mu_1, \mu_2$, (5) $\mu_1 - 0.1\sigma, \mu_2 + 0.1\sigma$ (6) $\mu_1 - 0.3\sigma, \mu_2 + 0.3\sigma$, (7) $\mu_1 + 0.6\sigma, \mu_2 - 0.6\sigma$. Note that the lines inside line 7 in the bottom figure are lines 1 to 6 in the top figure.
Figure 7.5: Sequential sampling criteria with different Type I and Type II errors. The means values are fixed as in Figure 7.2, and $\alpha, \beta$ are changed. The errors values based on the line number are: (1) $\alpha = 0.01, \beta = 0.1$, (2) $\alpha = 0.1, \beta = 0.01$, (3) $\alpha = 0.001, \beta = 0.001$, (4) $\alpha = 0.01, \beta = 0.001$, (5) $\alpha = 0.0001, \beta = 0.0001$, (6) $\alpha = 0.001, \beta = 0.001$, (7) $\alpha = 0.01, \beta = 0.01$. 
Figure 7.6: A simulated CA performance. Each triangle corresponds to an observation, and its value is $C_n$. The line that connects a set of triangles represents the segment that was used to reach a decision. The upper and the lower lines illustrate the decision threshold variations for that segment.
Chapter 8

Conclusion

8.1 Thesis Summary

This thesis investigated designing an online ECG biometric system with a large emphasis on feature extraction and continuous authentication. ECG signal is prone to noise and is a slow signal to acquire. Due to these properties of the signal, this thesis contributions can be divided into two subparts. First part of the thesis investigated the scenario of having small sample size data, while the other part studied having adequate amount of data.

ECG signal being one of the electrical biological signals that is generated from internal organ makes its acquisition interfere with other body generated electrical signals such as electromyogram. ECG signal is acquired by attaching electrodes to the body, and any movement in the sensors imposes noise to the ECG signal. Hence, we proposed a method for online outlier removal. This method was based on a one-class classifier concept. A dataset of 930 normal heartbeats were modeled using Gaussian mixture model, and the model that achieved the lowest equal error rate (EER) value was obtained with a Gaussian of two mixtures. This one-class classifier had an EER of \( \approx 12\% \) in the capability to separate normal from abnormal heartbeats. Such model can be used with
any subject without prior knowledge about the subject; hence, it is an online outlier removal system. This system achieved an EER of 5.94% in a 1,012 subjects UofTDB database when it was deployed in a biometric system, and it outperformed other online and offline outlier removal based systems. Despite the promising results achieved, it is worth mentioning that this system is aggressive and eliminated around 50% of the total number of heartbeats. It is ideal for biometric systems where data keep streaming such as a wearable device, however, using it in a system where the amount of data is limited raises the issue of small sample size.

Promising results of the ECG outlier removal with GMM along with the problem that may arise about small sample size led us to study the Gaussianity of ECG heartbeats and study the capability of synthesizing data to reduce data imbalance between the number of genuine and imposter observations. It is pertinent to mention that the segmented heartbeat had a duration of one second, and the data was sampled at 200Hz; hence, each segment had a size of 200 time-samples. Using Royston’s test and Sequential Forward Selection, a joint Gaussian distribution could be achieved with more than 20 time-samples. A joint Gaussian model was created, and synthesized data were generated by randomly drawing samples from the Gaussian distribution. When 20 real genuine heartbeats were used to generate Gaussian distribution and 200 synthesized heartbeats were generated, this designed biometric system achieved an EER of 6.71% in comparison to a minimum of 9.35% when data synthesis was not utilized (i.e. only 20 genuine observations were used). Biometric system with data synthesis outperformed several biometric systems which employed dimensionality reduction techniques. The EER of the biometric system with the synthesized data outperformed PCA by 3.21%, Probabilistic PCA by 6.76%, Isomap by 9.45%, Local Linear Embedding by 8.11%, and Laplacian by 6.73%. Lastly due to data imbalance, we used a parallel structure of bagging classifiers where each classifier was trained with same genuine observations but different imposter observations. This structure reduced classification variability among subjects. The true
acceptance rate standard deviation using parallel classifier was dropped from 6.52% to 1.94%.

Biometric system trained with ECG heartbeats time-samples suggested a viable approach to achieve a robust ECG biometrics, yet there is still some noise interference with ECG heartbeats that if tackled, a higher accuracy and robustness can be achieved. As a result, we desired a feature extraction method that can extract discriminative features from ECG heartbeats, is fast to extract, extracts temporal variation, tolerates various types of noise while having all that executed quickly. One-Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP) features type complies with all the properties mentioned. 1DMRLBP can tolerate shifting, scaling, misalignment, and quantization error noise. It encodes multiple time-samples into one value that captures the pattern of those time-samples. The 1DMRLBP has an extraction mechanism that extracts one feature vector from multiple observations. The subsequent classifier that is applied on the feature vector does not need to know how many observations were used in the construction of this feature vector. For example, the classifier can be trained with 1DMRLBP extracted from a single observation, and it can be used to classify 1DMRLBP feature vector extracted from \( n \) number of observations. This capability is crucial for the continuous authentication stage. 1DMRLBP features were experimented on two types of applications, biometrics and medical. In the biometrics application, the objective was to verify whether the claimed identity and the actual identity were the same. On the other hand, in the medical application, it was desired to classify a heartbeat whether it was healthy or unhealthy. The unhealthy heartbeats came from various types of heart irregularities. Comparison to state-of-the-art showed that a biometric system with 1DMRLBP achieved an EER of 7.89% while it was calculated to be 12.30% for AC/LDA based biometric system. The biometric system with 1DMRLBP without dimensionality reduction outperformed same system but with dimensionality reduction.

In applications where data kept collecting such as a wearable device, it was desired
to continuously find the identity of the person providing his/her signal. Conventionally, the designer decides the number of observations required to make a decision, then these observations are fused to obtain a single confidence value, and lastly, based on the confidence value, the system decides whether these observations belong to a subject or they do not. In this thesis, we desired a method that avoids setting up decision parameters and segment size, which is the number of observations needed to make a decision; hence, we used sequential sampling. Sequential sampling is a statistical method that decides based on past and current observations’ confidences the segment size and the decision thresholds. Sequential sampling has two criteria: accept criterion, and reject criterion, and if none of the criteria is met, more observations are needed. The distributions of the means of genuine and imposter classes are used to set up the criteria parameters. In sequential sampling, decision thresholds are adaptive and change in run-time. The only parameters that control the criteria are Type I and Type II errors. In this thesis, experiments were conducted on several parameters choices to examine the aggressiveness of the sequential sampling. Aggressiveness is defined by the separation width between the maximum confidence threshold before rejecting a hypothesis as being imposter and the minimum confidence threshold to accept a hypothesis that the claimed identity is true. The wider the gap, the more aggressive the sequential sampling is. There is a trade off, the bigger the separation between the lines (i.e. more aggressive), the higher the accuracy but more observations are needed. Therefore, the choice of parameters is application dependent. The proposed system could achieve 0.39% false rejection rate (FRR) and 1.57% false acceptance rate (FAR) with an average segment size of 32 observations. If a small number of observations was desired, then the proposed system achieved 1.93% FRR and 9.68% FAR with an average segment size of around 4 observations. Comparison to another state-of-the-art work that also used sequential sampling showed that the state-of-the-art work achieved FRR of 4.05% and FAR of 10.10% with an average of 35 observations, while it achieved FRR of 8.55% and FAR of 9.30% with an average of 6
observations. The proposed work performance surpassed another continuous authentication system that used a fixed-size segment and used decision making strategies including \textit{mean}, \textit{median}, 75\textsuperscript{th} percentile, 90\textsuperscript{th} percentile, 95\textsuperscript{th} percentile, and \textit{max}.

In the proposed continuous authentication system, the capability of 1DMRLBP in extracting one feature vector from multiple observations was utilized such that there was no need to deploy a fusion of observations in a segment to find a single classifier confidence. 1DMRLBP provides a single feature vector from multiple observations. To conclude the continuous authentication experiments, real-life simulation was conducted, and it well presented the phenomena of segment size change in the presence of noise.

It is worth mentioning that the contributions in this thesis aimed to design a complete online continuous authentication system, yet each contribution can be used independently in different applications.

\section*{8.2 Future Work}

This thesis was investigated for different applications yet only on ECG signals. The proposed types of outliers removal, feature extraction, continuous authentication with their properties can be studied on other types of one-dimensional signals. Some of the possible avenues to further advance these contributions revolve around:

\subsection*{8.2.1 Data Synthesis}

The number of available imposter observations is usually significantly larger than the number of available genuine observations. Research has acknowledged the imbalance and used a small subset of the imposter dataset in order to reduce it. Imbalanced data negatively affect performance. As demonstrated in this thesis, adding genuine data synthesis improved the results. We exploited the Gaussianity of ECG heartbeats, nevertheless, other approaches can be further researched which maybe oriented towards deep learning
to generate data. Preliminary results achieved encouraging results. Also, the parallel
classifier structure helped reducing accuracy deviation among subjects. Further investiga-
tion can be performed, specifically in training methods for such parallel structure.

8.2.2 1DMRLBP Parameters Optimization

One of the features of 1DMRLBP is the capability to capture local and global morphol-
ogy using different parameters as explained in Section 6.3. Different sets of parameters
achieved different results. In this thesis, we concentrated on the capability of capturing
multi-resolution in one-dimensional signal. However, it might be possible to set the pa-
rameters choice as an optimization problem and study the best suitable set of parameters
that extracts the highest distinctive features.

8.2.3 1DMRLBP Sparsity

The feature vector of 1DMRLBP is sparse; 90% of the feature vector had a value of zero
in the case when the parameters used constructed a 2,560 feature vector length. An
approach in image based binary patterns, called Uniform Local Binary Patterns [146],
events histograms for specific patterns. Patterns with a preset number of switches
between 0’s and 1’s. For example, U2 is Uniform Local Binary Patterns with 2 transitions
(e.g. 11000111 has one transition from 1 to 0 then one transition from 0 to 1, so there are
two transitions in total). All other patterns are grouped in one bin. With this strategy,
the work in [146] could reduce the feature vector size tremendously while keeping around
90% of the accuracy. The idea of capturing only specific patterns can be investigated,
also only patterns that occur frequently maybe considered for extraction. Unlike the
Uniform Local Binary Patterns, these patterns may have a specific template captured
from the training data rather than having a rule to extract them.
8.2.4 Continuous Authentication Model Adaptation

Diet intake, physical activity, diseases, and mood are some of the acts that destabilize ECG signal temporarily. Hence, a method to update the enrolled model to adapt to the changes might be needed. Model adaptation can be incorporated with continuous authentication. In [3], it was emphasized that there is no rule of thumb for choosing a specific time on how long the heart signal variability stays at its stage before reverting back to its previous stable stage. Consequently, it is difficult to decide on when to update the model. Model adaptation was incorporated with continuous authentication in [3] as well. In the context of this thesis, sequential sampling with two criteria was used for continuous authentication. The criteria split the decision into three regions: authentication region, continue region, and reject region. For model adaptation, one can update the saved model once an authenticate region is met.
Bibliography


