

ON ESTABLISHING PPG BIOMETRICS FOR HUMAN RECOGNITION: FEASIBILITY AND
VARIABILITY

by

Umang Natubhai Yadav

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
The Edward S. Rogers Sr. Department of Electrical & Computer Engineering
University of Toronto

Abstract

On Establishing PPG Biometrics for Human Recognition: Feasibility and Variability

Umang Natubhai Yadav

Master of Applied Science

The Edward S. Rogers Sr. Department of Electrical & Computer Engineering

University of Toronto

2018

This thesis presents a new biometric modality called Photoplethysmograph (PPG) for biometric authentication. PPG records the blood volume change from any part of the body with a combination of a light emitting diode and a photo diode. Therefore, it is the easiest to acquire amongst all medical biometric traits. Although PPG is a time-dependent physiological signal and alters with physical and psychological stress, most of the prior art has neglected these factors. This work presents the ‘BioSec.Lab PPG Database’ and establishes the PPG as biometrics through a comprehensive analysis on its distinctiveness and robustness under exercise, emotions, time-lapse, and noisy conditions. Performances under all variations mentioned above are assessed using various methodologies. Encouraging Equal Error Rates (EER) between 0.00% – 36.09% were achieved with relaxed states and under different variations from the three databases.

Acknowledgements

The first and foremost, I would like to give heartfelt thanks to my advisor Prof. Dimitrios Hatzinakos to give me an opportunity to learn and grow. During the last two years, I learned so many things intellectually, emotionally and professionally. This could not have been possible without the support, confidence, and freedom from my advisor in every way possible. I am forever grateful to him.

I would also like to thank my thesis committee members: Prof. Broucke, Prof. Kundur, and Prof. Plataniotis for their insightful questions and feedback.

I appreciate and acknowledge the financial support received from the NSERC, Department of ECE, SGS and Mitacs.

I am grateful to my colleagues and friends: Sherif, Jin, and Majid for the helpful discussions. I also thank my roommates and friends for making my time at UofT memorable.

Most significantly, I would like to thank my parents, sister and my late grandfather, to whom I owe my everything for their unconditional love and blessings.

Dedication

मातृ देवो भवः।
पितृ देवो भवः।

*to,
my parents*

Contents

Acknowledgements	iii
Dedication	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
List of Acronyms	x
1 Introduction	1
1.1 User Authentication	1
1.2 Photoplethysmograph (PPG) as Biometrics	2
1.2.1 Motivations	2
1.2.2 Challenges	3
1.3 Research Objectives and Contributions	4
1.4 Performance Measures	6
1.5 Outline of Thesis	7
2 Literature Review	9
2.1 PPG signal and its Applications	9
2.2 PPG Biometrics: Literature Survey	11
2.2.1 Fiducial Approach	11
2.2.2 Non-Fiducial Approach	15
2.3 Summary	17
3 PPG Database for Biometrics	19
3.1 Motivations for building a PPG Database	19
3.2 BioSec.Lab PPG Database	20
3.2.1 Recording Settings	20
3.2.2 Different Configurations	22

3.3	Other Databases	26
3.3.1	DEAP database	27
3.3.2	Capnobase Database	28
4	System Overview	29
4.1	Pre-processing	29
4.2	Feature Extraction	33
4.2.1	Continuous Wavelet Transform	34
4.2.2	One Dimensional Multi Resolution Local Binary Patterns (1DMRLBP) features	35
4.2.3	Auto-Correlation based features	38
4.3	Feature Reduction	39
4.3.1	Motivation for feature reduction	39
4.3.2	Linear Discriminant Analysis (LDA)	39
4.3.3	Small Sample Size Problem	40
4.3.4	Direct LDA	41
4.3.5	Motivation for Non-Linear Method	41
4.3.6	Regularized Kernel Discriminant Analysis (R-KDA)	42
4.3.7	Other feature reduction techniques	42
4.4	Classification	43
4.4.1	Support Vector Machine (SVM)	43
4.4.2	Addressing class-imbalance problem in training SVM	44
4.4.3	Random Forest based Classification	46
5	Empirical Results	48
5.1	Difficulties in engineering whole pipeline	48
5.2	Experiments with BioSec.Lab PPG Database	51
5.2.1	Results with Continuous Wavelet Transform (CWT)	52
5.2.2	Results with One Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP)	55
5.2.3	Results with Auto-Correlation (AC)	59
5.2.4	Comparison of results from all features	62
5.3	Results with Capnobase Database	65
5.4	Results with DEAP database	67
6	Conclusion and Future Work	69
6.1	Summary and Conclusion	69
6.2	Future Work	70
6.2.1	Improvements	70
6.2.2	Other Directions	71
A	ROC curves for BioSec.Lab Database	72
Bibliography		76

List of Tables

2.1	Summary of fiducial techniques	14
2.2	Summary of Non-Fiducial methods	17
3.1	Recording Device & other details	20
3.2	Details on different configurations of BioSec.Lab PPG Database	26
3.3	Details of DEAP and Capnobase database	27
5.1	Results with CWT as features for all configurations of BioSec.Lab PPG database	53
5.2	Results with 1DMRLBP and DLDA for all configurations of BioSec.Lab PPG database	56
5.3	Results with 1DMRLBP and R-KDA for all configurations of BioSec.Lab PPG database	57
5.4	Results with AC for all configurations of BioSec.Lab PPG database	60

List of Figures

1.1	Two modes of PPG recording	2
2.1	Sample PPG Signal	10
3.1	Plux pulse sensor and black Velcro strip used in this study	21
3.2	OpenSignal Desktop Application	21
3.3	Bitalino MCU and Placement of pulse sensor on fingertip with black Velcro strip	21
3.4	Example of PPG signals from single session for randomly selected subjects	22
3.5	Example of PPG signals before and after exercise for randomly selected subjects	23
3.6	Example of PPG signals from two different sessions in Short Time-lapse configuration	24
3.7	Example of PPG signals from three different sessions in Long Time-lapse configuration	24
3.8	Time difference between first and second session recordings, for Long Time-lapse configuration in BioSec.Lab PPG database	25
3.9	Time difference between first, second and third session recordings, for Long Time-lapse configuration in BioSec.Lab PPG database	25
4.1	System Flow Diagram	30
4.2	Frequency Spectrum Comparison of Raw and Filtered signal for subject 67 from in relax state from BioSec.Lab PPG dataset	31
4.3	Systolic peak and diastolic point detection in for subject ID 23 in 4 different sessions from DEAP dataset	32
4.4	PPG Segments	33
4.5	Comparison of Daubechies 5 mother wavelet and PPG Segment	35
4.6	CWT transform of PPG segment	36
4.7	Example of 1DMRLBP with $d = 4$ and $p = 5$	37
5.1	CWT performance comparison by taking best result each feature reduction technique	54
5.2	1DMRLBP performance comparison by taking best result each feature reduction technique	58
5.3	AC performance comparison by taking best result each feature reduction technique	61
5.4	Best Performance Comparison for all Features	62
5.5	Receiver Operating Curves comparisons for best results from all features	63
5.6	Comparison of results with Capnobase Database for different features	66

5.7	Capnibase ROC Comparisons	67
5.8	DEAP ROC Comparisons	67
5.9	Comparison of results with DEAP Database for different features	68
A.1	Receiver Operating Curves for CWT	73
A.2	Receiver Operating Curves for LBP	74
A.3	Receiver Operating Curves for AC	75

Acronyms

1DMRLBP One Dimensional Multi Resolution Local Binary Pattern

AC Auto Correlation

ANS Autonomic Nervous System

AUC Area Under the Curve

BASN Body Area Sensor Network

BioSec.Lab Biometrics Security Lab

CWT Continuous Wavelet Transform

DLDA Direct Linear Discriminant Analysis

DWT Discrete Wavelet Transform

ECG Electrocardiogram

EEG Electroencephalograph

EER Equal Error Rate

EOG Electro-Oculogram

FAR False Acceptance Rate

FRR False Rejection Rate

FT Fourier Transform

GDA Generalized Discriminant Analysis

HRV Heart Rate Variability

KKT Karush Kuhn Tucker

KPCA Kernel Principal Component Analysis

LBP Local Binary Pattern

LDA Linear Discriminant Analysis

LED Light Emitting Diode

LLE Locally Linear Embedding

LLRT Log Likelihood Ratio Test

MCU Micro Controller Unit

NNDD Nearest Neighbor Data Descriptor

OVR One Vs All

PCA Principal Component Analysis

PCG Phonocardiogram

PD Photo Diode

PPG Photoplethysmograph

PVC Premature Ventricular Contraction

RBF Radial Basis Function

R-KDA Regularized Kernel Discriminant Analysis

ROC Receiver Operating Characteristics

SDP Semi Definite Programming

SMOTE Synthetic Minority Oversampling Technique

SSS Small Sample Size

STFT Short Time Fourier Transform

SVDD Support Vector Data Descriptor

SVM Support Vector Machine

TAR True Acceptance Rate

TEOAE Transient Evoked Oto-Acoustic Emissions

THM Traube Hering Mayer

Chapter 1

Introduction

1.1 User Authentication

As the world is becoming more digitized and networked, a need for authentication is also rising. Authentication refers to the process of verifying the identity of a user, machine, or any other entity involved before granting them access to resources. Interestingly, user authentication can be carried out by three different ways that include: (1) what the user knows or knowledge based authentication; (2) what the user has or object based authentication; and (3) by identifying who the user is or using biometric characteristics [1]. Knowledge based authentication such as password or PIN relies on a user's memory. Supposedly, with a 4-digit PIN one has to make around 10000 guesses to intrude into the system. However, one can exploit a person's tendency to keep their PIN more memorable and can achieve the same result with fewer guesses. Therefore, long and complex passwords are insisted upon to make the system less vulnerable. But, doing so makes passwords more liable to forget and results in a tedious process. Object based authentication such as tokens or ID cards require a user to carry it all the time. In the event of losing it, security is compromised. Both knowledge based and token based authentication have their pros and cons. Therefore, a two-factor or a multi-factor authentication which comprises both is more popular [2].

On the other hand, a major advantage of using a biometric characteristic for authentication is that it does not have to be a secret or carried around, as it relies on a user's traits. In recent decades, use of biometrics has become more ubiquitous. Conventionally, physical characteristics such as the face, fingerprint, iris, palm print, etc. are used as biometrics. Major drawbacks of using these characteristics are that they are easy to steal, replicate, and do not have inherent liveness detection. For example, in an era of social media and surveillance, it is easier to get someone's face image. This face image can be used to perform presentation attack whereby the face recognition system can be fooled by this image [3]. Similarly, fingerprint and palm prints can be lifted from surfaces touched by a user. Iris images can be taken at a distance. These limitations have called upon many sorts of attacks that can be carried out by amateurs without requiring strong technical knowledge. Apart from physical traits, behavioral characteristics such as keystroke dynamics, voice, gait,

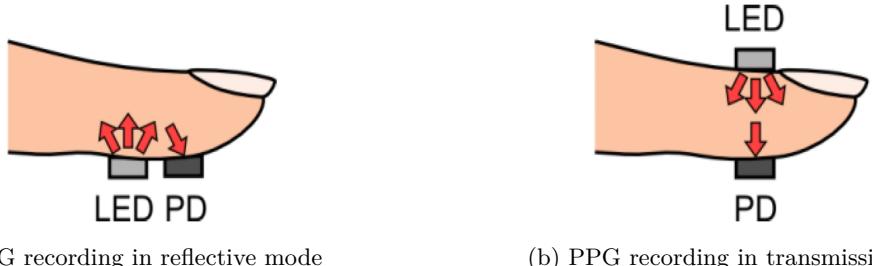


Figure 1.1: Two modes of PPG recording [4]

signature, and so on are also used as biometrics. Behaviors can also be imitated and therefore their use for authentication is more prevalent in forensics.

Limitations of physical and behavior traits have motivated researchers to instead use physiological traits for biometric authentication creating concept of ‘medical biometrics’. In medical biometrics signals such as Electrocardiogram (ECG), Electroencephalograph (EEG), Electro-Oculogram (EOG), Transient Evoked Oto-Acoustic Emissions (TEOAE), Phonocardiogram (PCG) etc. are used for user authentication. Being physiological signals, they provide inherent liveness detection. Additionally, they are difficult to record without a user’s knowledge, making it harder to steal. Contrary to physical traits, attacks on these systems cannot be carried out easily without technical depth. Despite these benefits, these signals require greater user co-operation. Additionally, they have less flexibility in choosing a recording site. For instance EEG and TEOAE can only be recorded from head and ears respectively. Except for TEOAE, all of them also need multiple electrodes for recording. Optionally, they require conductive gel for better contact with skin surface. On the other hand, TEOAE requires external stimulus into ears to stimulate oto-acoustic responses. All these factors result in less flexibility in device design, increased cost, and inhibit better user experience.

1.2 Photoplethysmograph (PPG) as Biometrics

1.2.1 Motivations

To address the aforementioned limitations of existing medical biometrics, this thesis will explore another physiological signal called Photoplethysmograph (PPG) for biometrics. Photoplethysmograph (PPG) is a non-invasive electro-optical method which measures the change in volume of blood flowing through the body part under testing. It reflects the pulsative actions of the arteries through the interaction of the oxygenized hemoglobin and photons. Hypothesis is that every person has a unique hemodynamics and cardiovascular system. Since, PPG captures these unique characteristics, it is investigated for biometric authentication herein.

Being a physiological signal, it carries the same advantages as other signals over physical traits. As shown in Fig. 1.1, PPG can be recorded with just a combination of Light Emitting Diode (LED) and Photo Diode (PD) from any part of the body, which provides greater flexibility for system design. Since, PPG only requires

LED and PD, it is extremely cost effective compared to all other biometric traits. Apart from authentication, PPG is widely used for many clinical applications, such as heart rate, oxygen saturation, respiration rate, blood pressure measurements and detecting peripheral vascular diseases. Since recording devices (LED/PD) are very small in size, they can be easily integrated into wearable devices and smart garments. For instance, PPG sensors are embedded in many commercial wearable devices such as Fitbit™ and Apple Watch for heart rate measurement. Hence, PPG also presents its application for user authentication in such devices. Compared to other medical signals, PPG does not require multiple electrodes, conductive gels, or any stimulus with advantages of greater flexibility in recording site, lesser cost, small form factor of sensors ultimately ensuing better user experience.

Replacing a photo diode in PPG recording with a camera makes Remote PPG or PPG imaging. Remote PPG records the color change in received light with a blood volume change. This contactless remote PPG has a high correlation with contact PPG. Remote PPG has become popular recently as a measure of anti-spoofing or liveness detection in face recognition systems [3], [5]. In the past, similar attempts have also been made with respect to fingerprints [6]–[8]. Being a cardiac signal, PPG is synchronized with ECG. Having a lower cost, small size of devices and ease of use, PPG also presents opportunity of multimodal biometric recognition system. PPG can also be utilized in Body Area Sensor Network (BASN) for generating secure keys for encrypting the messages along with authentication [9], [10].

1.2.2 Challenges

Apart from satisfying all operational requirements, to establish any new modality for biometrics, it should also satisfy the following properties [11]:

1. Universality: Biometric trait should be present in all human beings. PPG being a physiological signal, it is naturally present in all living human beings.
2. Distinctiveness: PPG is a very new biometric modality. Early work on PPG based biometric systems has shown promising results, demonstrating the distinctiveness of PPG among different individuals.
3. Permanence: Biometric traits should not change significantly over the time.
4. Conductibility: Data should be easy to acquire. As mentioned earlier, PPG is one the easiest to acquire among all biometric traits.

Moreover, a practical biometric system should also address following issues [11]:

1. Circumvention: It relates to how easily a biometric system can be fooled.
2. Performance: Given resources and operational environment, this characteristic relates to accuracy and the speed of a biometric system.
3. Acceptability: Given the cost, ease of use, social biases, and privacy concerns, acceptability relates to the extent to which people are willing to adopt it in their daily lives.

In order for PPG biometric system to realize its promising applications, it must satisfy all of the above mentioned criterion for a practical biometric system. While it can be argued that PPG biometrics can satisfy the acceptability requirement fairly, it is the *circumvention* and *performance* condition that pose major challenges for PPG biometrics. Since PPG recording needs sensory contact with a user's skin, it is difficult to steal without a user noticing. Another way could be generating PPG signal artificially. This would require knowledge of operation and dynamics of PPG and the system as a whole. Hence it can be argued that PPG biometric system, is not impossible but is certainly difficult to circumvent. Therefore, a major focus of this thesis will address the performance challenge.

In practice, PPG signal is generally impaired by many common noise sources during acquisition, such as motion artifacts, sensor movements, respiration, Premature Ventricular Contraction (PVC), ambient light, etc. Therefore, many of the early studies concluded that because of its noisy nature, a PPG biometric system should only be used under a controlled environment. More attention was given to alternative medical signal such as ECG. Moreover, physical or psychological stress increases blood pressure and heart rate. As a consequence, it changes the PPG signal. Empirical comparisons of PPG signal under rest condition and exercise show a considerable change in shape of PPG signal and its spectrum. Also different emotions influence the functioning of Autonomic Nervous System (ANS) and the heart. Spectral features derived from Heart Rate Variability (HRV) were shown to be useful in the emotional assessment but also with respect to the validating changes in PPG signals [12]. Such changes and noises can greatly affect the performance of PPG biometric system. Another major challenge is speed. Compared to traditional biometric traits, PPG is slow to acquire and requires at least 0.8–1.5 seconds to record a complete one period of PPG signal. In practice, just one period of PPG does not convey much information and cannot be used to produce satisfactory performance.

As a physiological signal, PPG can also change with routine activities, for example, a heavy diet, caffeine intake, and many other factors such as age, cholesterol level, red blood cells count etc. These changes serve as major impediments for PPG to satisfy *permanence* requirement.

1.3 Research Objectives and Contributions

Following the above discussion, research objectives for the work presented herein revolve around mainly three things. 1. Distinctiveness, 2. Permanence and 3. Performance. Permanence and performance also relate to the robustness of PPG as biometrics.

From the literature review presented in Chapter 2, it is clear that no large scale study has thus far been conducted on PPG biometrics. Most of the previous studies were only limited to relax state single session evaluation where training and testing data came from the same continuous recording in controlled environments. Moreover, the largest sample size used for any study was only 44 subjects. These studies also ignored the the effects of time and physical or psychological stress. Consequently, their results are unreliable for real life applications. Moreover, most of the studies directly utilized time domain signals or relied on different characteristic points of PPG signal for feature extractions. Hence, these features are more

vulnerable to noise and PPG signal variations with various factors.

Considering the aforementioned inadequacies, in this thesis we first study the effects of various factors on PPG signal and collect the largest biometrics database which has focused on PPG under various conditions (relaxed, exercise, after short and long time lapse). Secondly, we implement state of the art feature extraction techniques used in ECG biometrics together with different feature reduction techniques and classifiers to assess the distinctiveness and robustness of PPG biometrics. We also analyze advantages, limitations, and the effects of different parameters for each module in the proposed system.

Major contributions from this work are summarized as follows:

- In this work, we collected largest biometrics centric PPG database named ‘BioSec.Lab PPG Database’ from fingertips of participants using low cost PPG sensor under an environment which emulates real life. The database was collected under 4 different configurations. 1. Single session relax condition data from 84 subjects. 2. Across exercise from 40 subjects 3. Short time lapse from 55 subjects 4. Long time lapse from 37 subjects. Single session setting assisted in ascertaining the distinctiveness of PPG biometrics while others helped evaluate the robustness and permanence of PPG biometrics. Collected database is made public to encourage more work on PPG biometrics.
- This thesis has proposed and implemented a new PPG biometric system consisting of state of the art feature extraction techniques (Continuous Wavelet Transform (CWT), One Dimensional Multi Resolution Local Binary Pattern (1DMRLBP), Auto Correlation (AC)) with a focus on enhancing class separation using LDA based algorithms. Problem of Small Sample Size (SSS) during training was addressed using linear or kernel variants of LDA called (DLDA) and Regularized Kernel Discriminant Analysis (R-KDA).
- This thesis has presented a new online rule based algorithm for systolic peak and diastolic point detection that worked accurately across all variations in PPG signal. This algorithm can also be useful in clinical applications.
- Cost-sensitive framework was introduced to address issue of class imbalance during the training of Support Vector Machine (SVM) and Random Forest classifiers.
- Analyzed various limitations encountered during the training and tuning of the whole system from error and complexity perspective. This thesis has presented recommended parameters and other settings to achieve best results.
- This thesis has evaluated and compared the performance of various combinations of features, LDA, and classifiers on all configurations of BioSec.Lab PPG database to assess the distinctiveness and robustness of PPG biometrics. Promising results were achieved for all configurations. Results were also supported with insights. In addition, experiments with two other databases were also carried out to evaluate for the robustness towards respiration and emotional stress where better performance than state of the art was achieved.

This work is the first to extensively study the performance of PPG biometrics subjected to many variations

with a relatively large sample size.

Related Publication:

[13] U. Yadav, S. N. Abbas, and D. Hatzinakos, “Evaluation of PPG biometrics for authentication in different states”, in *2018 International Conference on Biometrics (ICB)*, 2018, pp. 277–282. DOI: 10.1109/ICB2018.2018.00049

1.4 Performance Measures

A biometrics system can operate in two modes. 1. Identification 2. Verification/Authentication. Prior to identification or verification, the enrollment of all subjects is carried out. The enrollment phase refers to the process of registering a new subject for future identification or verification purposes. Usually during enrollment, the biometrics data of a new subject is collected in controlled environments under various configurations. After which, the system is adapted or trained accordingly to carry out identification or verification in the future. Enrollment data is optionally stored and encrypted in the system database either in raw or processed form for future use.

Identification is a one to many ($1 : N$) classification problem. Given a query data in a $1 : N$ system, task is to identify best match among all pre-enrolled subjects in database. Identification could also be $1 : (N + 1)$ problem where instead of just finding the best match, additional condition is that matching score should be above pre-defined threshold. In this case $(N + 1)^{th}$ class refers to ‘no match found’.

Verification/Authentication is a binary classification problem in which the task is to differentiate between genuine and imposter subjects. A genuine subject refers to the legitimate user or positive class while imposter subject refers to fraudulent user or negative class. Compared to identification, here a true subject is known beforehand. The task here is to figure out whether query data in fact belongs to true subject or not.

In this thesis, the focus is only on verification based system. In this configuration, a subject can claim against an enrolled ID and based on the matching score and threshold, the system rejects or accepts the claim. In this setting has two types of errors which can occur. The system can falsely accept the impostor claim (False Acceptance Error) or can falsely reject a legitimate claim (False Rejection Error). To estimate these errors, a trained system is tested against the total number of G genuine trials and I imposter trials. For each trial, matching score is calculated and if it is above the threshold, then the trial is considered successful. Success of classification between imposter and genuine trials is measured by the number of True Positive (nTP), True Negative (nTN), False Positive (nFP), False Negative (nFN). The following quantities can be calculated from these numbers,

$$FAR = \frac{nFP}{I} \quad FRR = \frac{nFN}{G} \quad TAR = \frac{nTP}{G} \quad TRR = \frac{nTN}{I} \quad (1.1)$$

$$G + I = N \quad TAR = 1 - FRR \quad TRR = 1 - FAR \quad (1.2)$$

$$Accuracy = \frac{nTP + nTN}{G + I} \quad (1.3)$$

FAR and TAR stand for False and True Acceptance Rate while FRR and FRR stand for False and True Rejection Rate. N is total number trials.

It should additionally be noted that, in a class imbalanced dataset where $G \ll I$, accuracy would produce extremely optimistic results. For example, if $G = 10$ and $I = 1000$ then, classifying all trials as imposter trials would give $99 + \%$ accuracy even when $nTP = 0$. This means simply denying access to everyone would give high results, but that is not the purpose of verification. During testing, it is almost always the case that $G \ll I$, hence accuracy is avoided as performance metric in this thesis. On the other side TAR, TRR, FAR, and FRR separate the calculation of errors by using only one term in denominator, either G or I and measuring the numerator with respect to it. The issue of class-imbalance during training is further discussed later in Chapter 4.

For a biometrics application, one would have specific requirements on False Acceptance Rate or False Rejection Rate. For example, for high security application, this requirement could be FAR of 0.1% while for some other application it could be of 1%. To achieve a FAR of 0.1%, the threshold on matching score needs to be set very high to prevent imposters from getting access. But in doing so this could increase FRR, as genuine trials would also fall below high threshold because of intra-class variations. It can hence be seen as a trade-off. Decreasing FAR would increase FRR and vice-versa.

To capture this trade-off along with the different operating regions, Receiver Operating Characteristics (ROC) curve is used. ROC plots FRR vs FAR for several values of threshold providing performance comparisons under different operational requirements. Therefore, ROC curve and Equal Error Rate (EER) are adopted as performance metrics in this thesis. EER is a point on ROC where FAR=FRR. EER provides application independent general estimates on system performance. However, simply knowing EER is not reliable for selecting the best system. For example, it is possible that for one system, the value of EER is low while for the other, it is slightly higher. In this case, however, it is also possible that for specified allowance on FAR for an application, FRR is low for the second system while high for the first one. Hence, a ROC curve is needed to accompany EER. However, during this thesis, it was found that for almost all systems, ROC curves run parallel to one other. So in this case, a system with low EER is indeed the best system.

1.5 Outline of Thesis

The remainder of thesis is organized as follows.

Chapter 2 provides a detailed overview on the dynamics of PPG signal, how it is recorded, and factors influencing PPG signal and its applications. It also reviews related work on PPG based biometrics.

Chapter 3 presents ‘BioSec.Lab PPG database’ with details regarding its recording devices, the environment, and its configurations. It also includes a brief description about other two databases used during this

work.

Chapter 4 introduces a proposed system with details on each module design along with a review of each technique used. Following this, Chapter 5 discusses difficulties involved in training whole system. It reports and compares results for different variations in PPG signal.

Finally, Chapter 6 concludes the thesis with a summary and future directions.

Chapter 2

Literature Review

In this chapter, we first discuss PPG signal in detail with its components, various factors inducing variations in it and its applications. Afterwards, literature on PPG biometrics is reviewed, particularly in terms of methodologies, used database and evaluation framework, where it is also revealed that most of the previous studies lacked in accounting variations in PPG signal during its evaluation.

2.1 PPG signal and its Applications

Any system comprising PPG presents several challenges from a medical and engineering perspective. A PPG signal in general like any other physiological signal has two main components.

- AC component: It is the result of cardiac synchronous changes in blood volume with each heartbeat. AC part has a significant component between 1 – 2 Hz frequency. The AC part of the PPG signal can be defined in two phases. The anacrotic phase, defined as the rising edge of the pulse and concerned mainly with the systole phase of cardiac cycle. The second includes the catacrotic phase, defined as the falling edge of the pulse. The catacrotic phase is concerned with diastole phase of the cardiac cycle along with wave reflections from the periphery. A dicrotic notch is normally seen in the catacrotic phase of subjects with healthy arteries [14]. Figure 2.1 shows a sample PPG signal with fiducial points, such as systolic peak, dicrotic notch, and diastolic point.
- The DC component presents noises resulting from baseline wander, respiration, sympathetic nervous system activity, thermoregulation, vasomotor activity, vasoconstrictor waves, and Traube Hering Mayer (THM) waves. The DC part is mainly situated in lower frequency range on spectrum i.e. below 0.5 Hz [15].

Both AC and DC part of PPG are body site dependent [16]. Other noises that degrade the signal are ambient light, motion artifacts, improper sensor contacts, temperature, finger tremor, posture, coughing or changes in breathing patterns (arrhythmias, a deep gasp or yawn) [15] [17]. In addition to that, respiration modulates

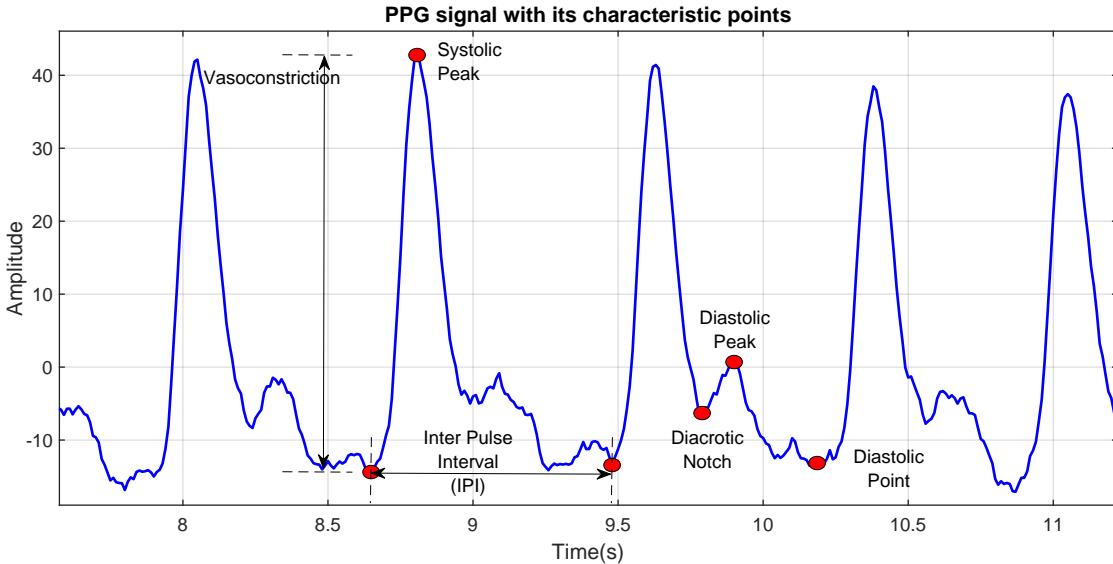


Figure 2.1: Sample PPG Signal with different fiducial points

the PPG signal introducing (a) Respiration Induced Frequency Variations (RIFV), (b) Respiration Induced Amplitude Variations (RIAV) and (c) Respiration Induced Intensity Variations (RIIV) [18]. These noises can be suppressed or reduced by proper isolation, thresholding, and filtering. From the device level too, there are some mechanical/electrical noises induced into the system such as power line noise, noises due to the improper spectral matching of LED and PD etc. Researchers are constantly improving sensor design to mitigate these issues [4].

A selection of body site for PPG recording is application dependent. Traditionally, PPG is recorded from either fingertip, earlobe, or toes. Amongst them, the fingertip is a good choice for biometric applications. There are two methods for the measuring the PPG signal. These include the transmission and reflective mode. In the transmission mode, as previously shown in Fig. 1.1, light is passed through fingertip or ear lobe and the response is recorded from the other side of the body part, hence named transmission mode. In reflection mode, light is incident on tissue using LED and the response is recorded with the PD from the same side.

Interaction of light with biological tissues is a complex process, which includes several phenomena, such as the transmission, (multiple) scattering, absorption, and fluorescence [19]. Hence, a selection of wavelength for PPG has been vastly studied in the biomedical community and is still an open area of research. Absorption of light in tissue happens mainly through water, melanin, oxy- and de-oxyhemoglobin. Red, near infrared, and green light is often used in PPG measurements. Red and near infrared light has lesser absorption providing greater penetration depth. This is one of the reasons why red and near infrared light is primarily used in the transmission mode of PPG. On the other hand, green light (530nm) is mainly used in reflective PPG. Studies have shown that green light is more robust to artifacts compared to red light, but has a shallower penetration depth compared to red and near infrared [20]. Apart from this, the orientation of red blood

cells and its count, blood volume, and blood vessel wall movement are amongst the key factors that affect the amount of light reflected by the tissue [15], [21], [22]. Most of the wearable devices that measures heart rate use green light as a source and reflective mode of PPG. For biometrics applications, reflective PPG with green light is more suitable as it is more tolerant to artifacts.

In 1937, Alrick Hertzman was the first to propose PPG as a technique to measure blood volume change [14]. Thereafter, PPG has found its place in a broad range of clinical applications. PPG is considered one of the most important signals which has transformed the health monitoring spaces. A few applications of PPG are listed below. More details on each of these applications can be found in [15].

1. Physiological monitoring : Blood oxygen saturation, heart rate, blood pressure, cardiac output, and respiration.
2. Vascular assessment: Arterial disease, arterial compliance and ageing, endothelial function, venous assessment, vasospastic conditions, microvascular blood flow and tissue viability.
3. Autonomic functions: Vasomotor function and thermoregulation, blood pressure, HRV, orthostatic intolerance, neurology and other cardiovascular variability assessments.

Replacing PD used in PPG sensor with video camera makes it ‘PPG Imaging’. In recent years, PPG imaging is being used as a remote contactless technique for monitoring and diagnosing ulcer formation, wound healing, skin blood flow and related rhythmic phenomena, oxygen saturation distribution within tissues, and so on [23]–[25].

2.2 PPG Biometrics: Literature Survey

Techniques for PPG based biometric recognition can be roughly classified into fiducial and non-fiducial based approaches. Fiducial methods rely solely on PPG signal itself and its different characteristic points for feature extraction while non-fiducial methods instead rely on some sort of feature extraction technique. In this survey, techniques where only systolic peak is detected for the purpose of segmenting PPG signal, are considered non-fiducial based approaches. Also, where temporal segment/signal is used directly for classification is also considered as non-fiducial. The survey includes a review of both identification and verification techniques.

2.2.1 Fiducial Approach

The earliest work on PPG biometrics was done by Gu et al. in 2003. Four fiducial features were extracted. These include the number of peaks, upward slope, downward slope, and the time interval between the bottom point and the first peak point. Matching was performed with Euclidean distance between the weighted features template. The weight was set for each feature using simple discriminant analysis. An accuracy of 94% was reported for 17 subjects. They concluded that a larger dataset under different conditions was required for further analysis and to put concept into practice [26].

Later in the same year continuing their work, Gu et al. presented fuzzy logic based decision making for authentication using gaussian membership function. With the same four fiducial features as previously on 17 subjects, an accuracy of 94% for single trial and 82.3% for two different trials were reported. Few false rejections were attributed to the poor quality of signal, respirations, and motion artifacts [27].

In 2007, Yao et al. demonstrated application of derivatives of PPG signals for identification purposes. They hypothesized that higher order derivatives of PPG might contain finer information compared to the original signal. Time intervals amongst maximum, minimum, and inflection points derived from first and second derivatives were used as discriminative attributes. A higher correlation amongst data from the same subject and lower correlation amongst different subjects were reported from a total of nine datasets recorded from three subjects only [28].

In 2009, Shi et al. investigated the possibility of non-contact reflectance PPG (NRPPG). NRPPG can be used in cases where contact with skin is not desirable, such as in the case of burned patients, neonatal monitoring, or patients with skin trauma. They set up vertical cavity surface-emitting LEDs (VCSEL) at certain distance and silicon PIN photo diode at 5 cm from tissue for recording of NRPPG. Simultaneous data of contact PPG and NRPPG from 22 different subject was recorded and signals were analyzed from both contact reflectance PPG and NRPPG using power spectral density, recurrence plot and bland altman plot. With the pearson correlation coefficient, a strong linear relation between different heart rate indices were shown. It was proposed that NRPPG can be used as an alternative to contact PPG for many applications, including biometrics. However, many issues such as poor signal to noise ratio, motion artifacts, tissue glare, and lack of dynamic range, were regarded as limitations [29].

In 2013, Bonissi et al. presented a preliminary study of PPG for continuous authentication. The method was very simple and based on maximum correlation based similarity score. The system was evaluated on two different settings. First on short time PPG dataset, where two minute data from 44 different subjects was divided into 20, 30, and 40 seconds of time segments. These time segments were used as a means for checking the effect of template update time on the performance of continuous authentication system and also to tune parameters for the recognition algorithm. EER of 5.29% was reported when updated every 40 seconds on two minute length of data. The second setting algorithm was tested on longer period of data, which was 15 minutes of data from 14 subjects. It was found that updating template after every 40 seconds gave a better performance of 9% EER. A degradation of performance with longer update time was attributed to physiological changes and insufficient durability of features [30].

In 2014, Kavsaoglu et al. made an effort towards fiducial PPG biometrics recognition system with feature ranking and feature selection methods. Forty fiducial features from raw PPG signal, its first derivative (FD), and the second derivative (SD) were also extracted. To boost the performance, Euclidean distance based class specific feature ranking method was employed. Several experiments varying number of features and value of k for template matching using kNN were performed. Fifteen period long PPG in two different durations from 30 subjects were recorded. The first PPG signals recorded from the subjects were evaluated as the first configuration. It gave identification of 90.44% when $k = 1$ and the first 25 features were used. In the second configuration, it was 94.44% with $k = 1$ and when only the first 20 features were used. Data from both configurations were combined to give way to a third configuration. An accuracy of 87.22% was

achieved, with $k = 1$ and first 15 features. Evaluation was done with leave-one out cross validation method [31].

In 2015, Lee et al. presented use of feed forward neural network for PPG identification. Twenty two fiducial features including waveform angle, area and inflection points were extracted for classification. The system was tested with data collected from 10 individuals. A false acceptance rate of 4.2% and false rejection rate of 3.7% was reported. Change in signal with exercise was also demonstrated. Additionally, the arm position and skin temperature were outlined as variables affecting PPG signal [32].

In 2015, Namini et al. presented a feature selection for PPG based biometric authentication. Peak detection for segmentation was done by detecting upward slope in pulse waves with threshold on duration of upward slope after passing signal through moving window integral. Thirty fiducial features were extracted using combinations of various features from raw signal, its first derivative and weighted derivative. Feature selection was employed with help of Forward Features Selection (FFS). For template matching, two comparison combinations (CC1 & CC2) schemes were proposed. Various tests were performed for classification by varying training size with kNN, fuzzy kNN, parzen window and Gaussian Mixture Model (GMM). Best EER of 2.17% was achieved with parzen window and CC2 on dataset of 30 subjects [33].

In [34], a preliminary study on twin identification using PPG biometrics was performed. Data was recorded in resting condition from three twins and one triplet where four subjects were female and five were male. Signals were processed with filtering and cycles were segmented with peak detection. Sixteen fiducial features were extracted. With naïve bayes classifier and radial basis function networks, an accuracy of 95.4% and 94.4% was respectively achieved. However promising, the study was limited to a total of nine subjects only and included only single sessions.

In 2016, Chakraborty et al. demonstrated the use of PPG in biometric recognition using three minute data recorded from 15 subjects. The method was based on 12 fiducial features obtained from PPG pulse. These features were divided into groups of two and standard deviation of these two groups of features were used to form the PPG template. Subsequently, LDA was applied. With the application of Euclidean distance, 100% accuracy was achieved [35].

In same year, Jindal et al. presented a deep learning based approach for PPG based identification using Deep Belief Net (DBN) and Restricted Boltzmann machine (RBM). Since, Heart Rate Variability (HRV) depends on factors such as weight, gender, etc., use of personalized models for each of the subjects was proposed. TROIKA dataset was used for evaluation which included data from 12 male subjects exercising under multiple stress levels. Data was preprocessed with moving average and low pass filter. PPG pulses were segmented using zero crossing and slope information in first derivative. Every segment was normalized in time length using auto regressive moving average. 11 features were extracted from time domain information and fiducial points. After finding total number of intrinsic clusters using GAP statistics, individuals were clustered into different subgroups using Partitioning Around Medoids (PAM) clustering to use personalized models. DBN with pre-trained stacked RBM was used for classification. Parameters for DBN were selected using grid search and gradients were estimated with contrastive divergence (CD-1). On average across all clusters, a classification accuracy of 96.1% was achieved with 10-fold cross validation [36].

Table 2.1: Summary of fiducial techniques

Paper	Features	Classification	# of subjects	Performance
Gu et al [26]	Number of peaks, upward slope, downward slope, time interval between bottom point and systolic peak	Simple Discriminant Analysis	17	94%
Gu et al [27]	same as above	Fuzzy logic	17	94%
Yao et al [28]	Time interval between extreme and inflection points	Correlation Coefficient	3	
Bonissi et al. [30]	Continuous authentication using mean of different number of time segments	Fusion of correlation measure amongst segments	44	5.29% EER
Kavasaoglu et al. [31]	40 fiducial features from FD and SD	Feature selection + kNN	30	94.44%
Lee et al. [32]	22 fiducial features	Feed Forward NN	10	3.7% FRR
Namini et al [33]	30 fiducial features	FFS + Parzen Window	30	2.17% EER
Nadzri et al [34]	16 fiducial features	Naive Bayes Classifier	9	95.4%
Chakraborty et al [35]	12 fiducial features	LDA + Euclidean Distance	15	100%
Jindal et al [36]	11 fiducial features	DBN + RBN	12	96.1%
Sarkar et al. [37]	GMM using 5 fiducial points	QDA	23	96%
Kamoi et al. [38]	9 Fiducial features	Random Forest classifier	15	8.8% EER

In 2016, Sarkar et al. examined variations in PPG biometrics with emotions using DEAP dataset. DEAP dataset consists of one minute data from 32 subjects under 40 different emotional states. Dynamical model given by McSharry was modified to represent PPG cycle. Signal was segmented into PPG cycles using custom peak detection technique. Dynamical model was used to normalize, align, and map the different lengths of PPG cycles into angular domain of $[0, 2\pi]$. Signals surrounding 5 fiducial points on PPG pulse were approximated with a sum of Gaussians during training. Several experiments were carried out varying training size and testing size using Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) as classifier. When evaluated with all available data from one subject without considering emotional state, 96% accuracy was achieved using QDA with 100 PPG cycles of training and 20 cycles for testing. However, when tested for emotional robustness, an authentication accuracy of 90.53% was obtained [37].

In study, by Kamoi et al. PPG signal was recorded while flicking hand wrist. Purpose of this study was to demonstrate use of PPG obtained from wearable devices such as smart wrist watch, while doing hand movement to replace activity or gesture based authentication. Peaks detected while doing this hand movement were used for authentication. Peak detection during motion was done after filtering using threshold on amplitude and time duration between two peaks. Nine fiducial features were extracted from these peaks. With respect to the dataset of 15 subjects, random forest classifier achieved EER of 8.8% when nine flicks or peaks were considered for classification [38].

2.2.2 Non-Fiducial Approach

In 2011, Spachos et al. examined feasibility of using PPG as biometric identifier using two different datasets with Linear Discriminant Analysis (LDA) as an automated way for identification. One of the datasets consisted of 14 subjects and produced an EER of 0.5%, while the other dataset consisted of 15 subjects and produced 25% EER. It was concluded that PPG biometrics system could produce better results under a controlled environment and with accurate sensors [39].

In 2012, Salanke et al. illustrated a feasibility of PPG biometrics with use of Kernel Principal Component Analysis (KPCA) as dimensionality reduction technique and mahalanobis distance as similarity matrix. Data from nine subjects were recorded in relaxed and stressed conditions at 37 Hz sampling rate for 60 seconds. Prior to template matching, peaks were detected in signal and signal was segmented around it. To address variability in segment length due to heart rate, all segments were scaled to 100 samples. High similarity was found between signals from relaxed condition while the distance between segments from the same subjects were found to have increased under stressed conditions. In the end, a satisfactory discriminability was shown among data from different subjects [40]. In yet another study by Salanke et al. (2013), researchers used Fourier analysis to reconstruct PPG signal impaired by motion artifacts. A semi-discrete decomposition was used to reduce the dimensionality of PPG feature vector. The same dataset was used with stressed conditions replaced by signal with motion artifacts. With Euclidean distance as the matching score, conclusions similar to earlier study were drawn [41].

In 2015, Jaafar et al. demonstrated use of Acceleration PPG (APG) (second derivative of PPG) for identification purposes. The system was tested on ten subjects having only four cycles of (APG). When tested against Bayesian network and kNN, Bayesian Network (BN) gave 97.5% identification rate as opposed to 90% with kNN. It was also shown that when using Raw PPG instead of APG, performance was poor, at 55% and 62.5% for BN and kNN. This paper concluded that APG had higher discrimination power than PPG [42].

In [43], time variability in PPG biometrics was assessed using decision tree models of Logistic Model Tree (LMT) and Functional Tree (FT). PPG signals from 5 subjects on two different days were extracted from MIMIC-II dataset. Signals were pre-processed using standard pipeline of filtering and segmentation. An identification rate of 96% was achieved using both classifiers.

Dias et al. presented exploratory use of PPG for continuous authentication. Evaluation method was entirely based on cross correlation of PPG time segments. Data was recorded while three participants were in a relaxed state and exercise state. For continuous authentication, data was collected in a relaxed state for one minute every hour for total of six hours. For exercise, data was collected for three minutes after exercise. Only three heart beats were used from each subject to form template gallery as it was proposed that due to higher variability in PPG, more than three heart beats degraded the results. It was also proposed that the alignment of PPG signals was necessary to achieve a high similarity. This was achieved with alignment of minima points in first derivatives of PPG or with Dynamic Time Warping (DTW). A correlation of 0.9 was achieved using only three heart beats for one subject. Additionally, a correlation of 0.82 and 0.25 was found when tested for signal recorded after seven days for the same subject and signal recorded just after exercise.

In the end for continuous authentication, high correlation was showed among data recorded at every hour for four hours. Moreover, for seven more subjects recorded separately for single session, a high correlation was indicated for the same subject [44].

In 2016, Choudhary et al. presented technique for systolic peak detection using Shannon Energy Envelope (SEE) for usage in biometrics application. Firstly, signal was filtered using Gaussian derivative filter to remove baseline wavers and high frequency noise. Before peak detection, the signal was squared and thresholding was done to suppress any low amplitude noise and to emphasize the amplitude of systolic peaks. Shannon Energy Envelop (SEE) was computed to reduce peak amplitude differences among systolic peaks. Smoothing filter was applied to suppress multiple peaks in SEE. Time instants of peaks were found by convolving SEE with Gaussian derivative kernel and finding negative zero crossing points. In addition, correction in detected time instances were done to prevent any error caused by smoothing function by selecting local maxima in fixed time window around detected peaks. Segments from 20s PPG signals were aligned around peaks with period and amplitude normalization and thereafter were used to construct one ensemble averaged pulsatile waveform for each individual in enrollment phase. For the selection of similarity score, experiments were carried out with Normalized Correlation Coefficients (NCC), Wavelet Distance (WDIST) and Wavelet based Weighted Percent Root Mean Square Difference (WWPRD). For the sample size of total 30 subjects, best EER of 0.29% was achieved with NCC. For the robustness analysis, artificial AWGN, power-line interference and baseline wander noises were added in PPG signals. The system achieved its best performance with NCC amongst all other similarity measures with FAR/FRR values in a range of 0.15-0.65% in many different noisy cases [45].

In a recent study, Karimian et al. (2017) presented non-fiducial based method for PPG based authentication. Signals were preprocessed using filtering and peaks were detected using modified Pan-Tompkin algorithm. Segmented PPG cycles were then normalized in amplitude and time. Features for each of the PPG cycles were extracted using Discrete Wavelet Transform (DWT) with coiflet as mother wavelet. To reduce correlation among features and to reduce the dimensionality of data, two-step process including Kolmogorov-Smirnov Correlation Based Filter (ksCBF) and Kernel PCA were applied. For experiments, Capnobase dataset was used which included eight minute data from 42 subjects. Classification was carried out using Support Vector machine, Nearest Neighbor Data Descriptor (NNDD) and Self Organizing Map (SOM). Best EER of 1.31% was achieved with NNDD using non-fiducial features from DWT and 9.53% when using fiducial features. Results concluded that a non-fiducial approach was superior to fiducial [46].

Later continuing their work, Karimian et al. studied the problem of human recognition using PPG. Preprocessing steps employed were similar to previous work. DWT using Daubechies wavelet of order four (db4) with four levels of decomposition was used for feature extraction. In this work, Genetic Algorithm (GA) was used to reduce dimensionality of feature vectors. Classification was carried out using SVM and Multilayer Perceptron (MLP). 100% accuracy was reported using non-fiducial features and 97.57% with fiducial features on same Capnobase dataset [47].

Later in same year, Yathav et al. presented new hardware to record both single lead ECG and PPG. Heart Rate Variability was extracted from ECG and PPG recorded from 20 subjects. HRV extracted from both PPG and ECG was proposed to be used as cross verification in biometrics systems. Comparative

Table 2.2: Summary of Non-Fiducial methods

Paper	Features	Classification	# Subjects	Performance
Spachos et al. [39]	Temporal Segments	LDA+Euclidean Distance	14	0.5% EER
			15	25% EER
Salanke et al. [41]	Temporal Segments	Kernel PCA	9	*
Salanke et al. [40]	Fourier Transform of time segments	Semi Discrete Decomposition + Euclidean distance	9	*
Jaafar et al. [42]	Temporal segments from APG	Bayesian Network + kNN	10	97.5%
Azam et al. [43]	Temporal segments	Logistic model tree & Functional Tree	5	96%
Choudhary et al. [45]	Temporal ensemble segment	Normalized Correlation Coefficients	30	0.29% EER
Karimian et al. [46]	DWT features	ksCBF + KPCA + NNDD	42	1.31% EER
Karimian et al. [47]	DWT features	Genetic Algorithm + SVM	42	100%
Sancho et al. [49]	Karhunen Loeve Transform	Euclidean Distance	42	4.8% EER

analysis showing a high similarity between HRV extracted from ECG and PPG [48].

Sancho et al. studied the time variability in PPG biometrics. Firstly, PPG signals were filtered and segmented around maximum peaks. Segmented signals were normalized along time and amplitude axis. Karhunen Loeve Transform (KLT) was as the feature extraction and Euclidean distance for matching. To assess time variability, three different scenarios were considered: Short-term, Mid-term and Long-term. In the short-term scenario, test signal was taken immediately after the train signal. In the mid-term, test signal was taken after seven minutes and in long-term, it was taken from the second day of recording. Capnobase dataset was used for short-term and mid-term evaluation and private dataset consisting of five subjects was used for long-term evaluation. Using 30 seconds of training and testing, Short and mid-term evaluation gave 4.8% and 9.5% EER on CapnoBase dataset. Long-term evaluation produced EER of 20% on private datasets. On small datasets, the results concluded degradation in performance with time [49].

2.3 Summary

PPG signal like any other physiological signal is affected by many factors that depend on each individual's characteristics and is also influenced by many noise elements. Since, it depends on an individual's heart functioning and cell constituents at time of recording, PPG signal shows variability with time, age, and any daily activities in general.

A key takeaway from literature survey on PPG biometrics is that despite having known that PPG alters with many factors, prior studies completely ignored these variations, with the exception of a few. Moreover, the largest sample size of any study was just limited to 44 subjects and in fact most of the studies had a sample size less than 40. The survey included both identification and verification methods and therefore, if only verification methods are counted then it can be seen that a smaller amount of work has been done in

the past on PPG-based verification systems.

Another observation can be made is that almost half of the studies, used fiducial features that are more prone to errors and noise. On the other hand, many non-fiducial methods in the survey used temporal segments directly without any feature extraction. Hence, we advocate that there is certainly a need for investigating different feature extraction techniques to improve performance especially when PPG shows greater intra-subject variability with time and under physical or when a participant is under psychological stress.

Chapter 3

PPG Database for Biometrics

In this chapter, we present BioSec.Lab PPG database recorded at the University of Toronto. This chapter starts with motivating ideas behind recording this database and then carries on to provide detailed description of different recording scenarios, recording protocol, difficulties faced during recording, and general observations. Along with BioSec.Lab PPG database, this chapter also includes details on the other two databases used during this thesis' research work.

3.1 Motivations for building a PPG Database

As discussed in Chapter 1, to establish any new modality as biometric identity, one must prove its uniqueness and permanence while satisfying all other requirements. Being a physiological signal, PPG changes with physical stress, mental stress, and over time. Evaluating PPG as a biometrics modality subject to all these scenarios is extremely essential.

Despite having known these criterion, most of the studies in the past were just limited to single session performance evaluation where training and testing signal were recorded continuously in one sitting. However satisfactory results might be from such studies, but they simply could not be considered as enough evidence for PPG as a ‘good’ biometrics modality. In addition, the largest sample size for any study was limited to 44 subjects [30]. From the survey in Chapter 2, it can be seen that there was only one paper which assessed permanence in terms of time-variability [49]. In that study also, time gap between two sessions was just limited to single day, which was less pragmatic setting. There was only one paper which evaluated the robustness of PPG under physical stress–exercise [36]. However, this was only was also limited to 12 subjects. There was only one paper which focused on the cross evaluation of PPG under different emotions, however, it was also limited to just 23 subjects [37].

The first paper on PPG biometrics is found to be published as early as 2003 [26]. Hence, despite this idea having been around for around 15 years now, no large scale study has ever been conducted. There are

Table 3.1: Recording Device & other details

Device	Databse specifics & Details
Recording device	Bitalino Micro-Controller Unit with Plux 520 nm green PPG sensor
Interface	Bluetooth Interface with OpenSignals Desktop Application
Sampling rate	100 Hz
Resolution	10 bits per sample
Recording Duration	3 minute
Participants Age	19-35 years

various reasons for the absence of a large scale study, but one obvious reason is a lack a public biometric database. There are many databases available publicly on PPG, but all of them are single session recordings focused on other applications such as heart rate measurement, blood pressure measurement, or respiratory rate measurement. There was no public database which was built with a focus on biometrics. Hence to stimulate more research on PPG biometrics, to put more emphasis on evaluating permanence property of PPG, to steer research outside of single sessions is proof of concept studies, it was a necessity to build one database. In the next section we introduce BioSec.Lab PPG database focused on addressing these issues.

3.2 BioSec.Lab PPG Database

BioSec.Lab PPG database contains PPG signals recorded from fingertip of participants under 4 different configurations namely a single session, across exercise, short time-lapse and long-time lapse. Data was collected in Biometrics Security Lab (BioSec.Lab) at the University of Toronto (UofT) under the ethics protocol # 33788. All participants were UofT students, and had no prior cardiac problems and with age range between 19–35 years. Duration of every recording was fixed to three minutes. Database does not contain any personal information from any subjects. Data is anonymized and participants are only referred by randomized subject ID. This database is made available to other researchers with Material Transfer Agreement (MTA) as per UofT’s policy [50].

3.2.1 Recording Settings

- **Recording device:** For data collection, Plux pulse sensor was used [51]. It has both LED and PD on the same side, hence it records a reflective type of PPG. LED operates on green light with a wavelength of 520nm. Sensor connects to Bitalino Micro Controller Unit (MCU) [52], which sends real time data via Bluetooth to OpenSignal desktop application [53]. As shown in Fig. 3.3, Pulse sensor is placed on the fingertip of each subject with the help of a black Velcro strip. The Velcro strip help maintain good contact with skin surface and provides isolation with ambient light. Sampling rate of this device was fixed at 100 Hz. It should be noted that, we did not rely on medical-grade or precision devices for recording. All devices used here are portable, small in size, cheaper in price, and lightweight enough to carry anywhere around. Further to this, the mobile version of Opensignals application is also available, facilitating PPG recording



Figure 3.1: Plux pulse sensor and black Velcro strip used in this study

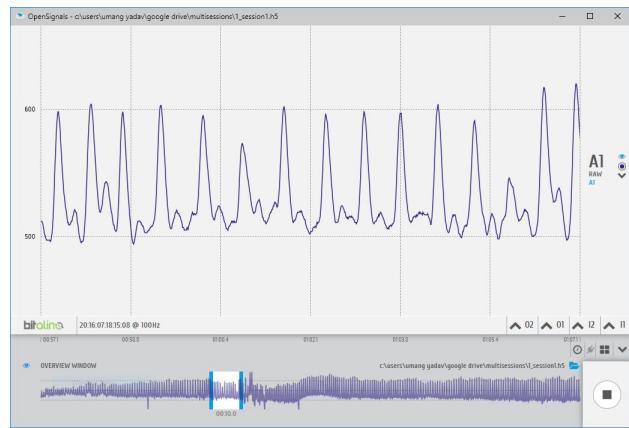


Figure 3.2: OpenSignal Desktop Application

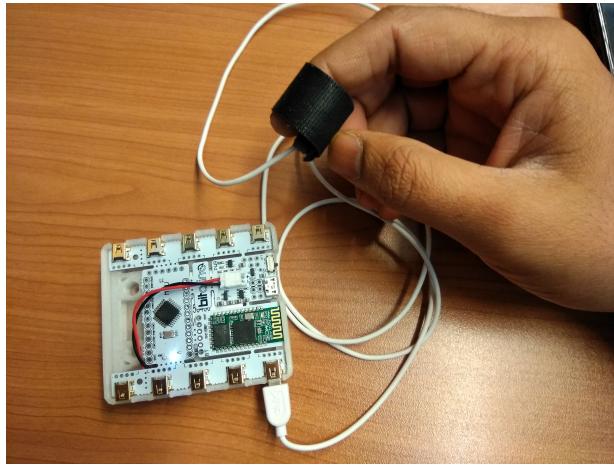


Figure 3.3: Bitalino MCU and Placement of pulse sensor on fingertip with black Velcro strip

on the go anywhere via Bluetooth. Hence, these devices indeed emulate practical recording devices such as the smart-watch.

- **Followed protocol before each recording:** An open call for participation was circulated to invite as many people as possible. Common protocol followed during each recording is as follows. Prior to recording sessions, all participants were asked to sit down on a chair in a suitable position and relax for a while to achieve a normal heart rate. Participants were briefed about objective of this study and its importance. Any questions from them were answered and an informed consent form was signed for data collection. Then, Plux sensor was placed on either their left or right index finger-tip, irrespective of which fingertip was used in earlier sessions in multi-session configurations. Proper sensor contact was ensured by adjusting Velcro strip. It was found that a tight grip with Velcro strip resulted in poor signals. Therefore in all cases, it was made sure that Velcro strip provided moderate grip.

- **Unconstrained subject:** To minimize motion artifacts, subjects were requested to avoid any hand

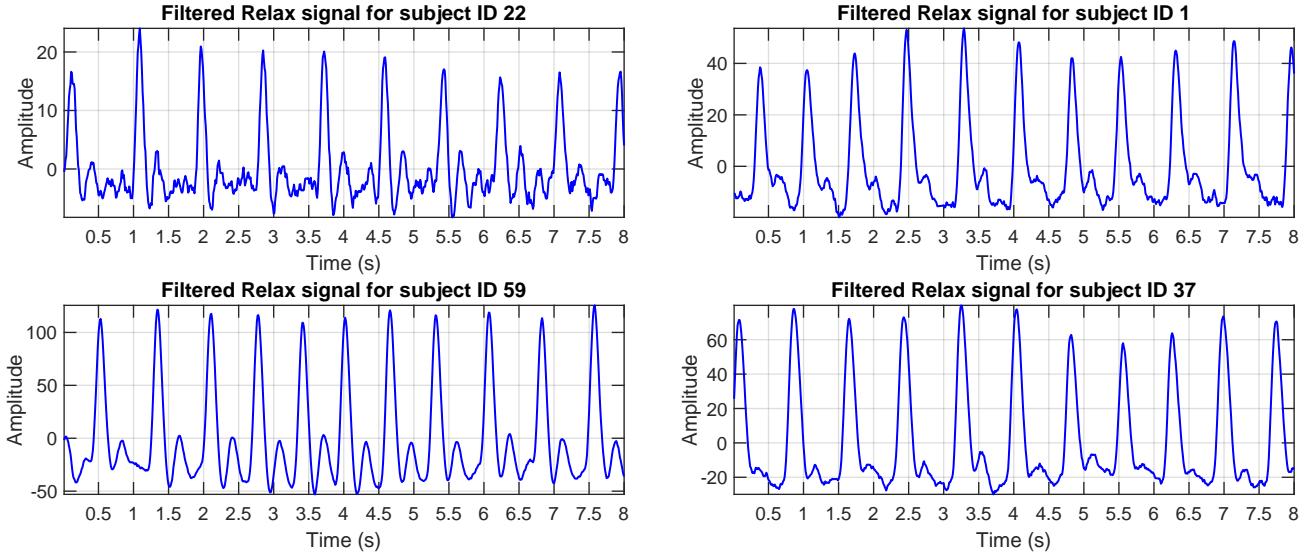


Figure 3.4: Example of PPG signals from single session for randomly selected subjects

movements on which the sensor was placed. Except for that, no other constraints were put on subjects' behavior during recording. Participants were allowed to talk, move their legs, and use their cellphones with their free hand. They were allowed to follow any breathing pattern for example, relax, sleep, meditate, yawn, etc.

- **Recording Environment:** All recordings took place in a normal office environment where intended application is imagined to be. Hence, signals were recorded in an environment where many people were talking, eating, typing on the keyboard and so on, which could simulate usual distraction. The office where the study was conducted had many computers and power-intensive servers running, which could cause power line interference. Thus the database is indeed collected in more practical settings.

3.2.2 Different Configurations

- **Single Session:** Single session refers to PPG signal recorded in one sitting in a relaxed state where it was recorded continuously for three minutes. This is the most common setting to identify usability of any medical biometrics modality. BioSec.Lab PPG database contains single session data from 84 subjects. Prior to recording, same settings as in section 3.2.1 were followed. This is the largest known sample size so far for any PPG biometrics database. Fig. 3.4 shows PPG signals for four subjects recorded in this configuration.

- **Across Exercise:** To evaluate the effect of physical stress, signals were recorded after doing heavy exercise. Subjects were asked to perform any cardio exercise for sufficiently long time to achieve higher heart rate. One example of exercise is climbing up and down the stairs fast. Signals were recorded for three minutes immediately after exercise while heart rate was very high. Recordings from 40 subjects are available under exercise. For assessing the effect of exercise on the performance of biometrics system, relax condition 3 minute signals for all 40 subjects recorded before exercise are also accompanied with exercise recordings.

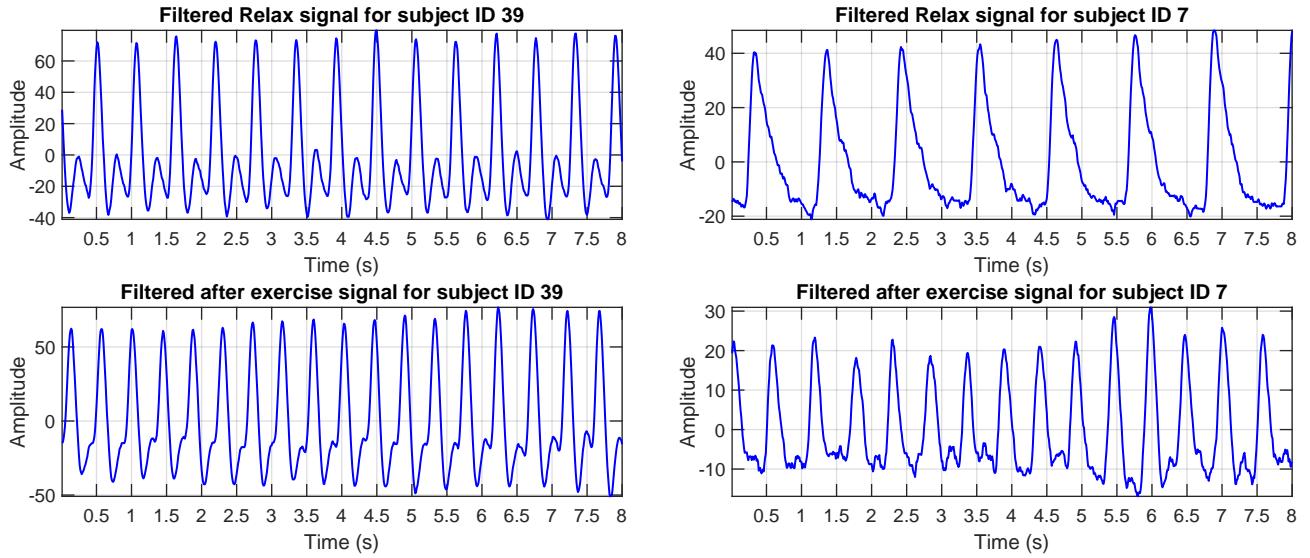


Figure 3.5: Example of PPG signals before and after exercise for randomly selected subjects

Fig. 3.5 shows an example of PPG signals for two subjects recorded before and after exercise where change in HR and signal shape can be noticed.

- **Short Time-lapse:** Short time-lapse configuration is multi-session recordings where signals were recorded with comparatively short period of time gap between two sessions. Time-difference between two recordings was fixed to be a minimum of 30 minutes on same day. The motive behind recording in this configuration is twofold. First, to evaluate permanence property of PPG biometrics subjected to short time difference and daily activities. Second, to evaluate with single session recording is less practical and results could be misleading as training and testing is done on continuously recorded signal. Short time-lapse configuration provides more practical setting with break between training and test samples.

Because of short time gaps in these recordings, participants were invited to come for a free lunch meal as an incentive to stay for two sessions. First session recordings were carried out before lunch and second session after lunch. There were also free caffeinated soft-drinks along with meal as supplements. Heavy consumption of diet has an effect of slowing down heart rate while caffeine has the opposite effect. Hence, it is expected that along with time variance, PPG signals would also be affected by diet and drinks. The purpose here is not to evaluate the effects of diet, but in general, the effect of day-to-day activities and time variability on PPG. In addition, not all participants consumed meal or drinks. Therefore, this configuration comprises good mix of variations. In total data from 55 subjects are available in this configuration. Because of only one recording device, accommodating all these participants was a challenge and signals were recorded one subject after another hurriedly as all participants were students who came for recordings between short break in their classes. Thus, these recordings also simulate more practical and uncontrolled environments. Fig. 3.6 shows an example of PPG signals from two sessions recorded in a short time-lapse configuration.

- **Long Time-lapse:** For this configuration, the time difference between two recordings was fixed to be at least of 14 days hence named long time-lapse. Subjects participated in single session were invited for

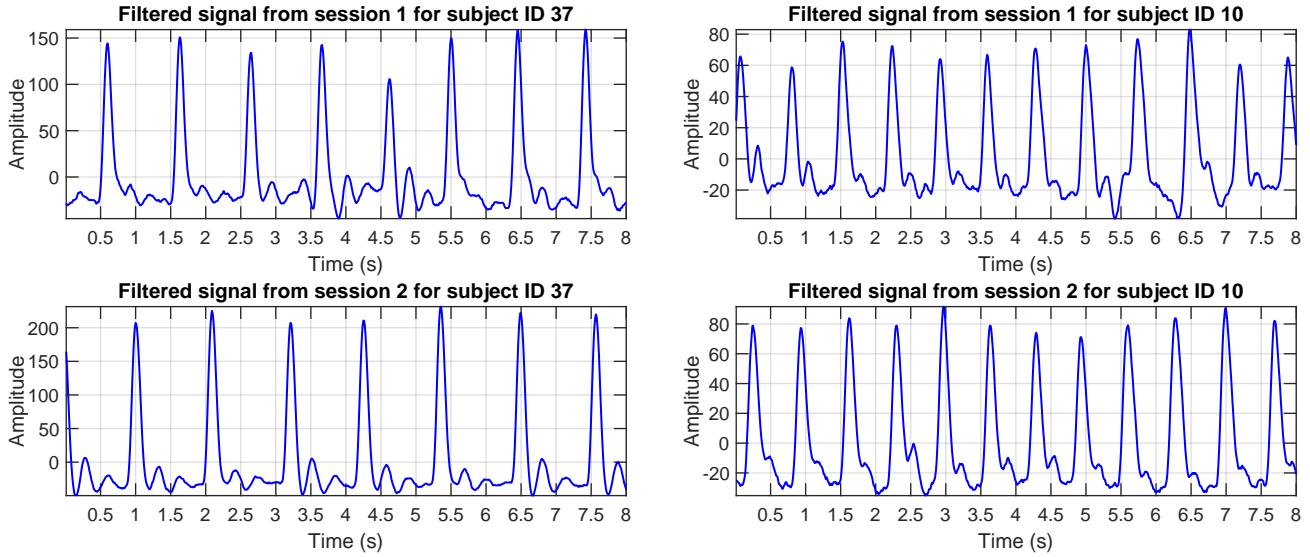


Figure 3.6: Example of PPG signals from two different sessions in Short Time-lapse configuration

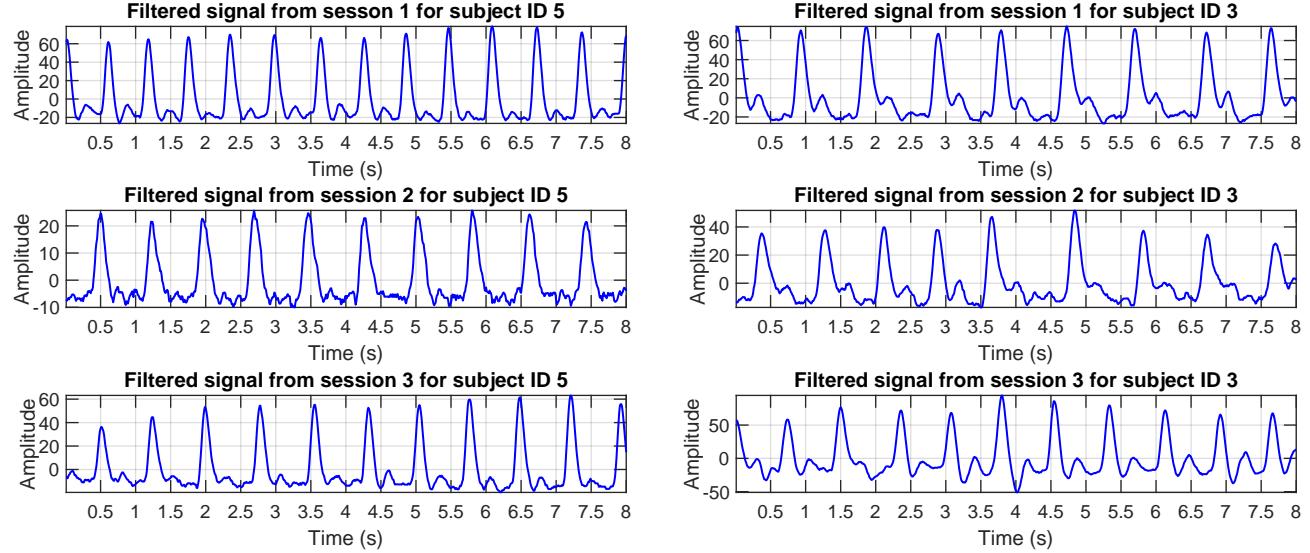


Figure 3.7: Example of PPG signals from three different sessions in Long Time-lapse configuration

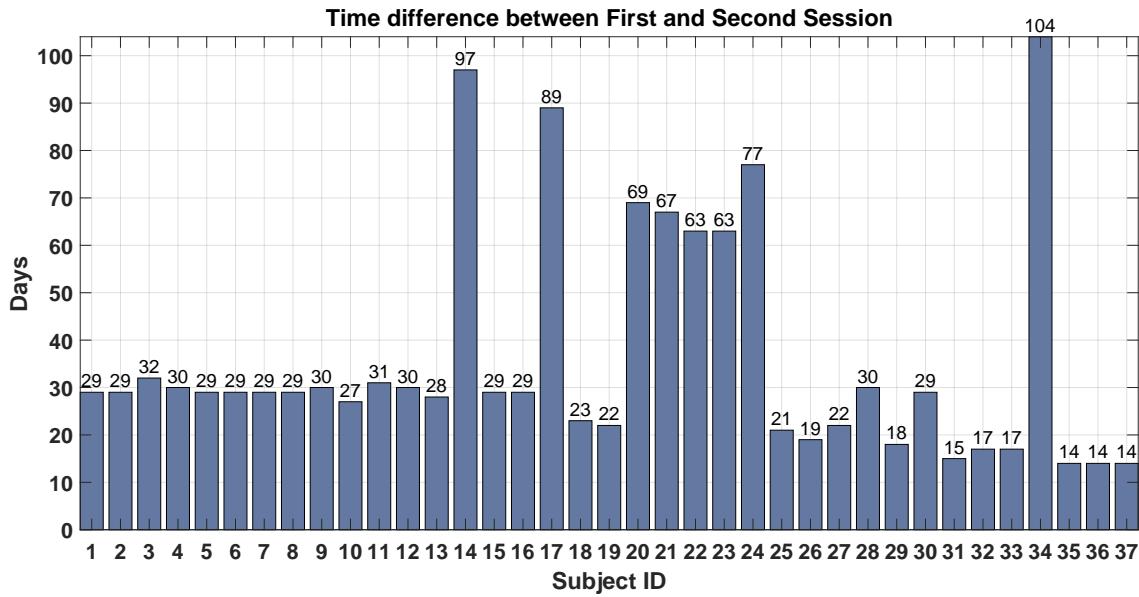


Figure 3.8: Time difference between first and second session recordings, for Long Time-lapse configuration in BioSec.Lab PPG database

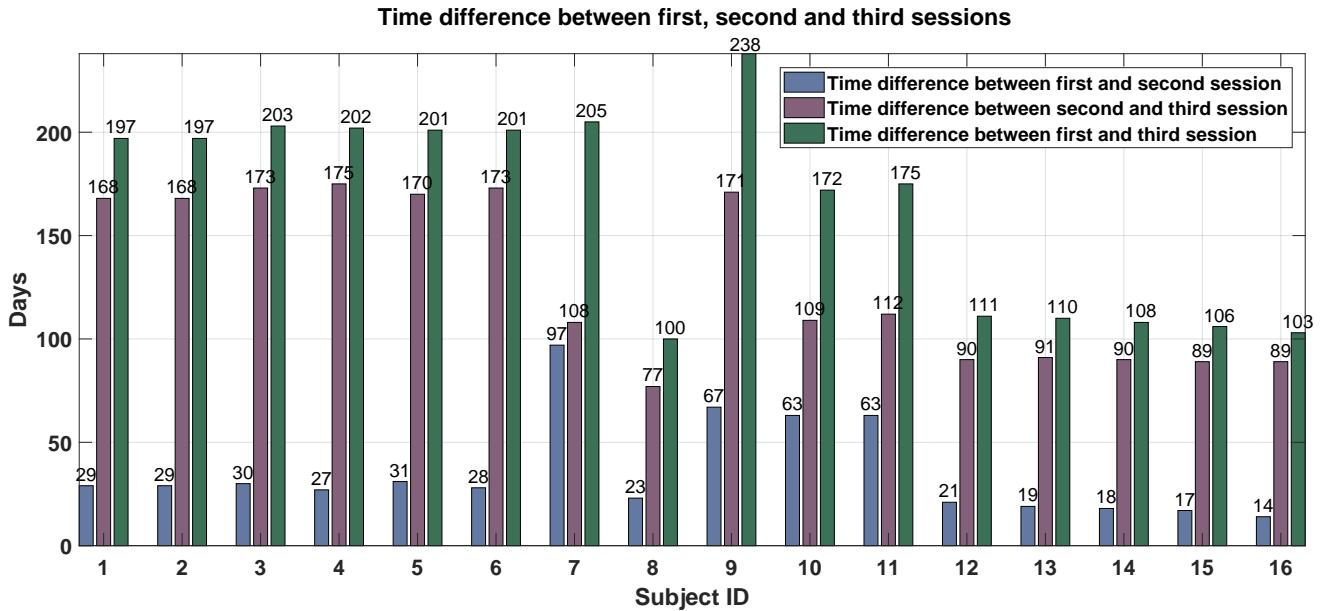


Figure 3.9: Time difference between first, second and third session recordings, for Long Time-lapse configuration in BioSec.Lab PPG database

Table 3.2: Details on different configurations of BioSec.Lab PPG Database

Configuration	#Subjects	Details
Single Session	84	Relaxed state.
Across Exercise	40	Signals from both relaxed and after exercise.
Short Time-lapse	55	Two sessions with a minimum of 30 minutes time difference on same day.
Long Time-lapse	37	Two sessions with a minimum of 14 days time difference in between. On average 36 days of time difference in between.
	16	Three sessions with a minimum of 14 days between any two sessions. On average 36 days of time difference between the first and second sessions. On average 128 days of time difference between the second and third sessions.

the second and third session of recording. Same recording settings as with single sessions were followed. In total, 37 subjects participated for second session. Although, the minimum time difference was set to 14 days but the time difference between the first and second session for all subjects was 36 days on average. Exact details on time difference for each subject is given in Fig. 3.8. Out of these 37 subjects, 16 subjects also came for third recording session. The average time difference between the second and third session was 128 days. For experimental convenience, recordings of participants with only two sessions of data and subject with all three sessions recordings were stored in separate files. Also, as shown in Fig. 3.8 and 3.9, these two files have different time-difference characteristics. From these same figures, it can be observed that these long time-lapse configuration contains wide range of time-variations. For few subjects, it is as large as 6 months. Fig. 3.7 shows an example of PPG signals from three sessions where it can be seen that signals from different sessions are not exactly same.

3.3 Other Databases

Apart from BioSec.Lab PPG database, two other databases were also used during this thesis. These two databases have been used by few other papers. Thus, it is imperative to use the same databases for comparison of our method with other papers. Secondly, these two databases were recorded with a focus on different applications and have been recorded in different environments. Assessing our system performance on differently recorded databases would also give a better estimate on the robustness of PPG biometrics. DEAP database was recorded under variety of elicited emotions while Capnibase database was recorded with an aim of finding respiration rate from PPG signals [54][18]. Emotions certainly affect participant's heart rate and in turn PPG. Hence, it is also essential to evaluate performance of PPG biometrics under emotional stress. Respiration induces noises in PPG signal which is harder to eliminate with filtering. Hence, it is also important to assess performance with a database that has been recorded especially for the respiration-rate measurement.

Table 3.3: Details of DEAP and Capnibase database

DEAP Database		Capnibase database details	
Number of subjects	32	Number of subjects	42
Number of sessions for each subject	40	Sampling Rate	300 Hz
Duration of each recording	1 minute	Length of each recording	8 minute
Sampling Rate	128 Hz	Age of pediatric subject	8.7 ± 5.5 years
Age with mean age 26.9	19-37 years	Weight of pediatric subject	35.6 ± 23.3 kg
Gender Ratio	50 % male : 50% female	Age of adult subjects	47.2 ± 9.7 years
		Weight of adult subjects	73.5 ± 24.2 kg

3.3.1 DEAP Database [54]

DEAP database is one of the benchmark databases in affective computing. It consists of one-minute PPG signal recorded along with various peripheral physiological signals for 40 different kinds of emotions which covered whole valence-arousal plane. Database consists of such recordings from 32 different subjects. It was formed with an aim to find correlation between various physiological signals while the subject was going through different kinds of emotions. Authors also presented single trial classification for emotions such as arousal, valence, and like/dislike using modalities of EEG, peripheral physiological signals, and multimedia content analysis on DEAP database in [54].

Emotions were stimulated using music videos. A pool of 120 such music videos were created using two different mechanisms such that it covered all four quadrants of valence-arousal space (LVHA-low valence high arousal, LVLA-low valence low arousal, HVLA-high valence low arousal, HVHA-high valence high arousal). Furthermore, only one-minute segments of these 120 videos which included maximum emotional content was selected using affective highlight algorithm. To select 40 music videos out of 120, which had the highest quantity of emotional content, an online subject annotation was carried out where participants rated each video on a discrete 9-point scale of valence, arousal and dominance. For each of the 120 videos, 14-16 such ratings were available. To maximize the strength of elicited emotions, a final selection of 40 videos was done such that it had the strongest ratings and lowest variations in ratings. On normalized valence-arousal plane, these videos lied on extreme corners of each quadrants and 10 videos were selected from each of the quadrants guaranteeing maximum emotional elicitation. Each experiment started with two-minute baseline recording. Then for each of the 40 recordings, the following steps were followed [54].

1. Two seconds of screen displaying current trial number to inform participant of their progress.
2. Five seconds baseline recording.
3. A one minute display of a music video.

After 20 trials, short break was given to participants. Since, this database contained signals where emotion elicitation was attempted to be maximized, it was used to evaluate the emotional robustness of our system.

3.3.2 Capnibase Database

Capnibase database was collected by [18] with purpose of development of improved respiratory monitoring algorithm for children and adults. 42 cases were randomly selected by [18] from a larger collection of physiological signals collected during elective surgery and routine anesthesia. It consists of single session, 8-minute finger PPG data recorded from 29 pediatric and 13 adults during controlled breathing. This database was used in few recent paper on PPG biometrics [46], [47], [49]. Our system was also evaluated on this database to compare performance with other papers on same database. Being focused on measuring the respiration rate under controlled environment, this database also contains comparatively larger respiration induced noise in PPG. Hence, it was also necessary to assess PPG biometrics performance under such conditions.

Chapter 4

System Overview

In this chapter, we present non-fiducial biometric authentication system based on PPG beats. The system flow is similar to many other biometric authentication systems. The flow diagram of our system is presented in Fig. 4.1. System modules can be grouped into pre-processing, template creation, and classification. In this chapter, we discuss all these modules in great detail.

4.1 Pre-processing

- **Filtering:** After acquiring raw PPG signal through sensor, it is first passed through the 4th order Butter worth band pass filter with cut-off frequencies 0.5 and 40 Hz. Filtering is needed for the removal of power line interference, baseline wavers. It also helps suppress motion artifacts and respiration induced noises. Fig. 4.2 shows an example of frequency spectrum of raw and filtered signal, where it can be seen that filtering process effectively removes the DC bias or any high frequency noise. Filtered signal was re-scaled on amplitude to have dynamic range of one using Eq. 4.1.

$$\mathbf{x}_{norm} = \frac{\mathbf{x} - min(\mathbf{x})}{max(\mathbf{x}) - min(\mathbf{x})} \quad (4.1)$$

- **Systolic Peak Detection:** Successful application of heart beats or PPG beats based non-fiducial pattern recognition techniques depends heavily on reliable peak detection. For example, false peaks could inject noise in training data or missing peaks could reduce available training data. Unlike ECG, there is no standard universally accepted algorithm for systolic peak detection in PPG. In practice, most of the local maxima finding algorithms work reasonably well with noise-free PPG signal. In noisy or motion artifacts ridden signal, such a naïve approach fails. In this work, an algorithm based on sliding window was developed for systolic peak detection and diastolic point detection. Sliding window combined with rules based on participants' heart rate and peak separation produced very accurate locations even in the presence of artifacts. This algorithm can also be employed for continuously streaming data, as it only requires samples

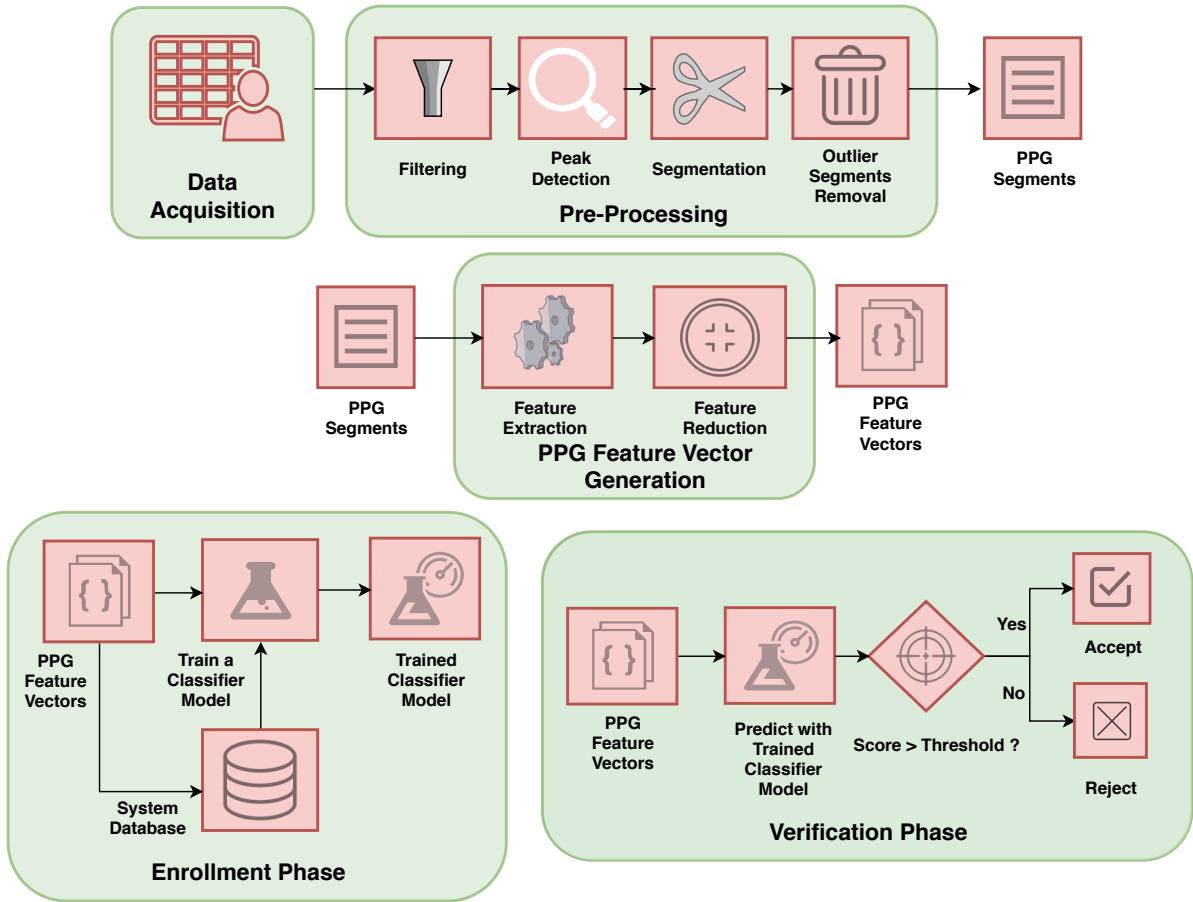


Figure 4.1: System Flow Diagram

from small window. Compared to many other techniques, our algorithm is extremely simple and requires less computations.

Pseudo-code for our algorithm is given in Alg. 1. For systolic peak detection firstly, one second window of PPG signal is taken. Windowed signal is then re-scaled using Eq. 4.1 and squared to emphasize amplitude differences between peaks and other points. Squaring is popular trick employed before peak detection. It is also the first step of the Pan-Tompkin algorithm used for peak detection in ECG [55]. For this squared windowed signal, a local maxima is considered as a peak, if its amplitude is higher than the preset threshold counted from reference valley point (Peak Prominence property). The reference valley point is the highest local minimum point that occurs just before or after a peak. The threshold on amplitude difference was set empirically to 0.2. Additionally, to remove false peaks, a threshold of 0.7 was applied on peak amplitude height and minimum distance between any two detected peaks was set to $0.4 \times \text{Sampling Frequency}$. Window is then slid by 0.5 second and the procedure is repeated for a whole signal. For diastolic point detection, global minima is found inside window of 0.3 second before every systolic peak. All parameters and thresholds were set empirically to give best performance. In practice, they can be varied for other applications or other signals.

Fig. 4.3 shows this method in action across different sessions for DEAP dataset. It should be noted that

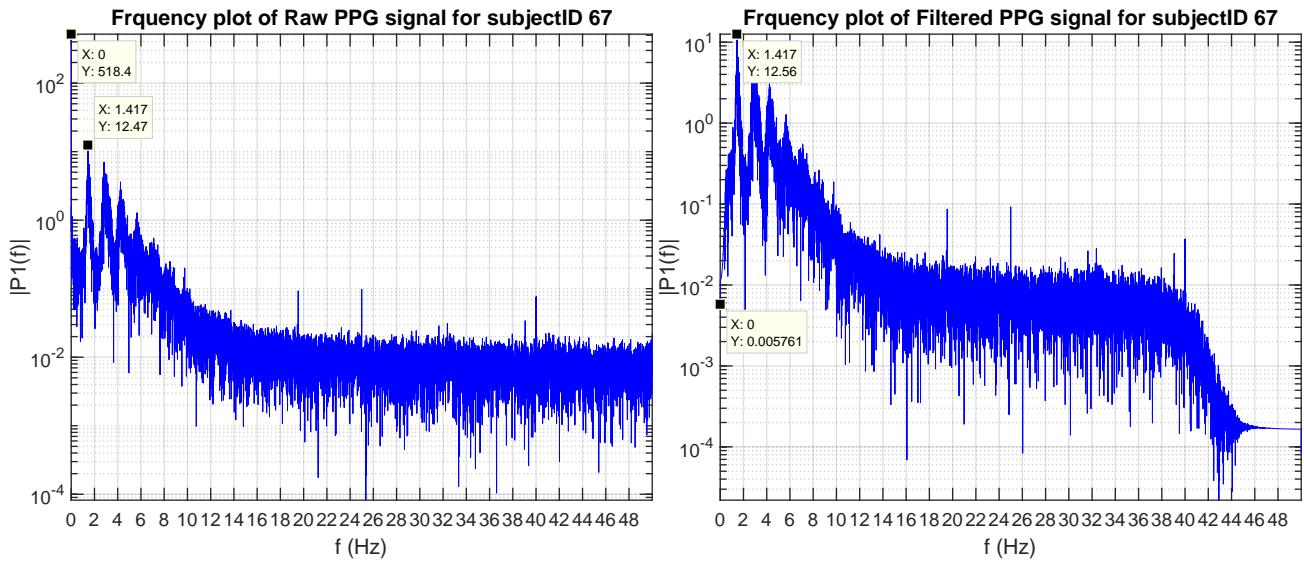


Figure 4.2: Frequency Spectrum Comparison of Raw and Filtered signal for subject 67 from in relax state from BioSec.Lab PPG dataset

the amplitude of PPG varies a lot across different sessions, hence simple threshold on amplitude cannot be relied upon for peak detection. Moreover, diastolic peak occurs just a few samples after systolic peak in a PPG signal. Hence, when amplitude varies, simple rule based peak detection would falsely detect diastolic peak as systolic peak.

Accurate measurements of systolic and diastolic peak locations are extremely important in many of the PPG's clinical monitoring applications e.g. cuff less arterial blood pressure measurement, heart rate variability, respiration rate measurement, sleep apnea etc. [18], [56]–[58]. Therefore, the biomedical community has put good efforts on accurately detecting these fiducial points. The majority of these techniques follow the same algorithmic patterns developed for the ECG. In biomedical literature, many of these fiducial point detection techniques use complex analysis based on power spectrum, slope of the curve, first/second differentiation of PPG and inflection points associated with it, moving average filters, adaptive threshold etc. Few researchers have also extended the idea for remote-PPG and developed fiducial point detection algorithms for remote-monitoring systems [59] [60].

In practice, we find that our algorithm accurately detects systolic peaks and diastolic points even for the after exercise and noisy signals. Since, focus of this thesis was not developing a fiducial points detection technique, comparative analysis with other methods in bio-medical literature were not carried out. However, we strongly believe that this algorithm has very good performance and could be of great benefit for clinical applications. In biometrics literature, most of the papers have excluded details on peak detection algorithm. For the rest of the papers, very few details are given on peak detection for anyone to reproduce or reuse the method. Additionally, most of the papers have been just focused on single session relax condition data. Performance under noisy or after exercise data is unknown. Since, we carried out extensive study, we find that it is important to provide all details on peak detection to facilitate any future studies.

Algorithm 1 Systolic Peak and Diastolic Point Detection Algorithm for PPG

Input: PPG signal \mathbf{x} , window length w , window shift Δ , Sampling Frequency f_s
Output: Systolic peaks and Diastolic points locations

```

Calculate total number of valid windows:  $l = 1 + \lfloor \frac{k-w}{\Delta} \rfloor$ 
Let  $s_i$  represent  $i^{th}$  PPG window
sysPeaks  $\leftarrow [ ]$  ▷ store systolic peak locations here
for  $i \leftarrow 1$  to  $l$  do
    rescale  $s_i$  using Eq. 4.1
     $s_i \leftarrow s_i^2$  ▷ take square of window
    windowPeaks  $\leftarrow$  find peaks in  $s_i$  with peak prominence  $\geq 0.2$  and amplitude  $\geq 0.7$ 
        and minimum distance of  $0.4 \times f_s$  with each other
    append windowPeaks to sysPeaks
end for
sysPeaks  $\leftarrow$  remove peaks in sysPeaks with distance between them  $< 0.4 \times f_s$  ▷ to remove false peaks or
twice detected peaks because of  $\Delta$  shift on boundaries of PPG windows

diaPoints  $\leftarrow [ ]$  ▷ store diastolic point locations here
minScan  $\leftarrow 0.3 \times f_s - 1$  ▷ Length of scanning window for diastolic point,  $0.3 \times f_s$  is set empirically
for  $j \leftarrow 1$  to  $\text{length}(\text{sysPeaks})$  do
    z  $\leftarrow$  sysPeaks( $j$ ) ▷ take  $j^{th}$  peak
    windowMin  $\leftarrow$  find global minima point in window from  $\mathbf{x}(z - \text{minScan})$  to  $\mathbf{x}(z)$ 
    append windowMin to diaPoints
end for

Return sysPeaks and diaPoints

```

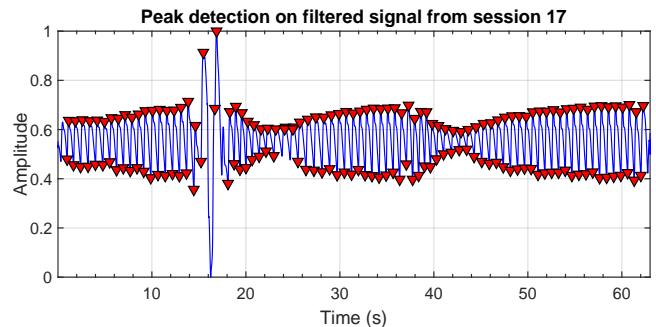
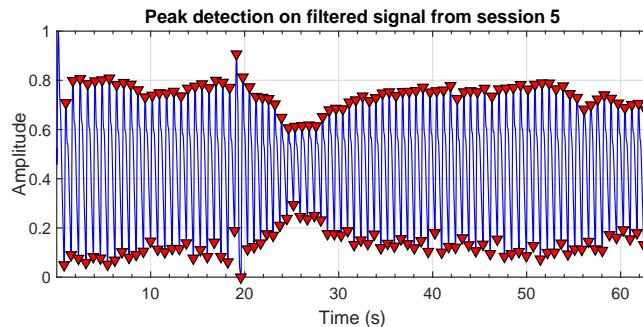
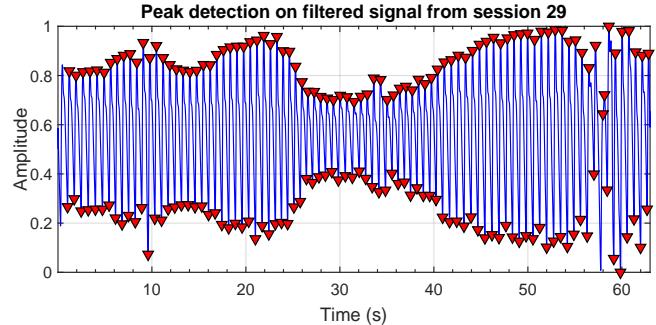
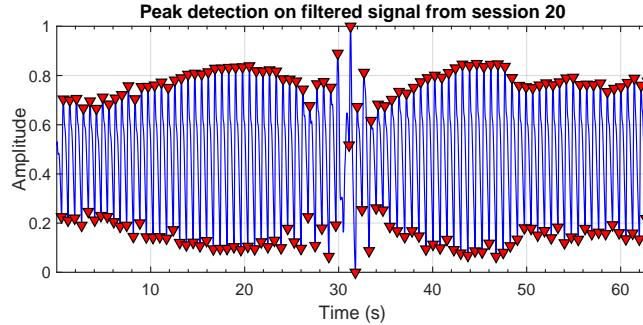


Figure 4.3: Systolic peak and diastolic point detection in for subject ID 23 in 4 different sessions from DEAP dataset

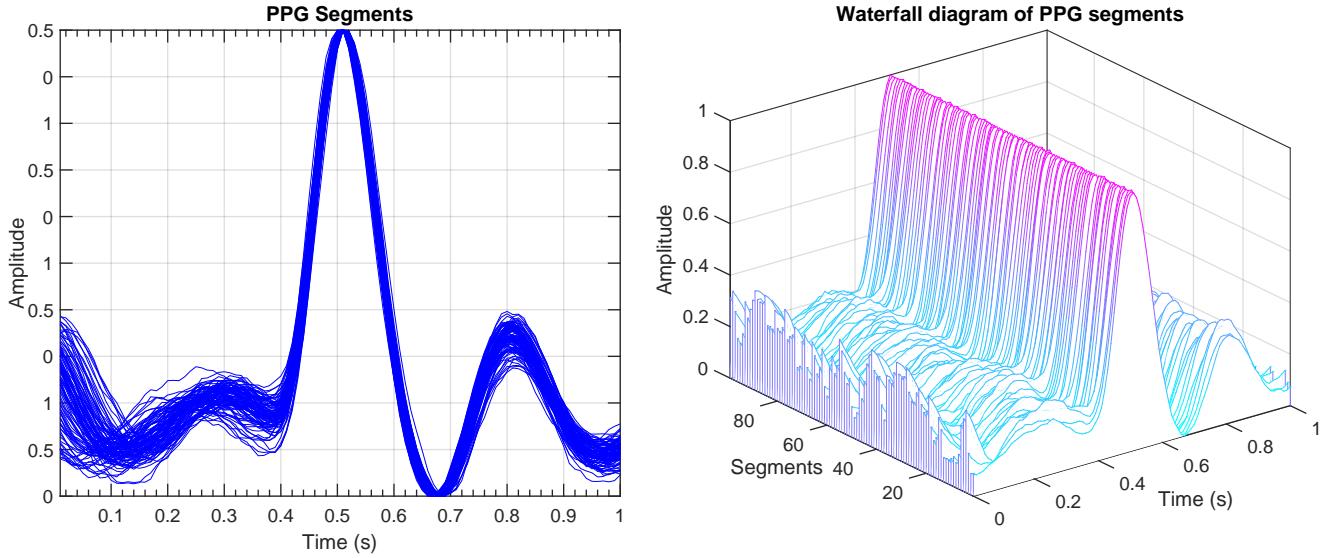


Figure 4.4: PPG Segments

- **Segmentation:** For the features extraction, signal was segmented around detected systolic peaks. The length of each segments was set to 1 second with 0.5s before peak and 0.5s after peak.
- **Outlier Segments Removal:** Any segment having Euclidean distance of more than pre-set threshold from median was considered an outlier and was removed from consideration. The threshold was empirically set such that amplitude variation of 0.2 on average was allowed on each sample of the segment. Cleaned PPG segments were then sent for feature extraction. Fig. 4.4 shows an example of PPG segments. It should be noted that a very stringent threshold could reject many segments making the training and testing time longer. On the other side, we would like to allow for intra-class variations in PPG segments in various conditions. Having a higher threshold could allow noisy segments to pass through, deteriorating performance. Hence, it is a trade-off.

4.2 Feature Extraction

Raw PPG time signal on its own is not very useful for biometrics. It is necessary to emphasize idiosyncrasies in PPG signal for its biometrics application and for that it is needed to extract informative features. From literature survey in Chapter 2, it is evident that most of the prior work on PPG biometrics directly relied on temporal segments or fiducial points as features. Therefore, it was also imperative for us to try out different feature extraction techniques. During this work, Continuous Wavelet Transform (CWT), One-Dimensional Multi-Resolution Local Binary Patterns (1DMRLBP), and Auto Correlation (AC) were investigated as features to improve performance under variations. A selection of these feature mainly stemmed from their good performance with ECG based systems. Moreover, each of these features carries distinct advantages. Following sections provide a review on these features along with its benefits and motivations for using them

for PPG in particular.

4.2.1 Continuous Wavelet Transform

Continuous Wavelet Transform (CWT) was selected for feature extraction. Compared to Fourier Transform (FT) and Short Time Fourier Transform (STFT), CWT is better at localizing time and frequency components of a signal. At its core, all these techniques try to represent time domain signals in terms of some basis function using inner products as similarity metric.

In Fourier Transform, this basis function is a complex exponential $e^{j\omega t}$. FT of a signal given by $F(\omega) = \int_{-\infty}^{\infty} x(t).e^{-j\omega t} dt$. Higher value of $F(\omega)$ at any particular value of ω represents higher similarity between signal and basis function at a value of ω . In FT, summation is taken over $-\infty$ to ∞ , so time information is completely lost. In other words, we can not determine which frequency component is present at what time. On the other hand, frequency information is accurately obtained.

To alleviate time resolution problem in STFT, fourier transform is taken over a short window of a $STFT(\tau, \omega) = \int_{-\infty}^{\infty} x(t).w(t - \tau).e^{-j\omega t} dt$. Here, window length is prefixed before taking STFT and it determines the time-frequency resolution trade off. For example, low frequencies require longer window length (larger support) to be correctly identified compared to high frequency. In practice, frequency behavior of a signal cannot always be known beforehand and fixed length window is not useful. In such cases, STFT performs poorly as it gives frequency bands rather than correct components resulting in poor frequency resolution. To overcome this issue, the concept of wavelets was introduced.

Continuous Wavelet Transform (CWT) of a signal is given by Eq. 4.2. In our case, $x(t)$ is a PPG segment. $\psi(t)$ is the mother wavelet, which acts as both a window and basis function, which is then stretched (scaled) or compressed (dilated) by parameter $a > 0$ to change window length. Here, shorter windows facilitate detecting presence of high frequency components and longer windows help with determining low frequencies. Since low frequency requires longer window time, time resolution is poor at low frequencies. On the other hand, time resolution is better at high frequencies, but it has poor frequency resolution. This wavelet is then translated across signal to generate time-frequency representation. By such translation and scaling, the mother wavelet separates mixed frequency components of given PPG segment. Here also the wavelet coefficient $WT_x(a, b)$ can be interpreted as inner product similarity metric.

$$WT_x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \cdot \psi^* \left(\frac{t-b}{a} \right) dt \quad (4.2)$$

Despite time-frequency resolution trade off, by nature CWT is very good at detecting abrupt changes, representing oscillations and revealing time-frequency structure of a signal. That was the primary reason for using CWT as features for PPG biometrics. Compared to Discrete Wavelet Transform (DWT), CWT is taken on finer values of scale (a). In DWT, scale is only changed in power of base 2 i.e. $a = 2, 2^2, 2^4 \dots$ and translation(τ) is also done in proportion of scale. Additionally, DWT is not shift-invariant. Hence, DWT is

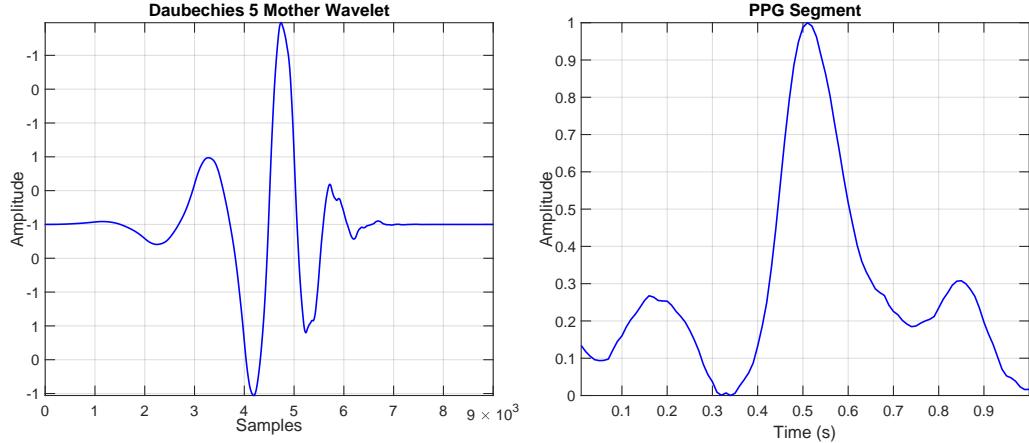


Figure 4.5: Comparison of Daubechies 5 mother wavelet and PPG Segment

only good when sparse or compressed representation is needed. On the other side, since CWT takes finer value of scale, it is a redundant transformation as coefficients do not change much with scale on finer level. For remedy, a number of scales can be controlled or sampled appropriately while extracting CWT features.

Heuristically, a mother wavelet is chosen such that it best resembles signal under inspection. In this work, Daubechies wavelet (db5) was chosen as mother wavelet. Fig. 4.5 shows db5 wavelet and a PPG segment for comparison. For feature vector generation, the signal was first decomposed from 1 to 128 scales to retain high frequencies at lower scales and low frequencies at higher scale. Then, coefficients from 32 scales were sampled at an interval of 4 from 1 to 128. This was done to keep dimensions of feature vector small and to reduce redundancy in representation. Based on the sampling interval of PPG segment, these scales roughly covered frequencies from 66 Hz to 0.5 Hz. Fig. 4.6 shows an example of CWT, where it can be seen that at lower scales shape of signal, onset, and offset of peaks are captured while at higher scales long-term or overall characteristics of signal is captured. Final feature vector for each PPG segment was of length $32 \times 100 = 3200$.

4.2.2 One Dimensional Multi Resolution Local Binary Patterns (1DMRLBP) features

One Dimensional Multi Resolution Local Binary Patterns (1DMRLBP) was first proposed by Louis et al. for ECG biometrics [61]. It was inspired and adapted for one-dimensional signals and in particular for biometrics application from commonly used Local Binary Pattern (LBP) features in Computer Vision.

4.2.2.1 Motivations for using 1DMRLBP for PPG

1DMRLBP features were proposed with attention on continuous authentication systems where continuous data stream or abundant data is available, for example in a smart watch or other wearable devices. Louis et

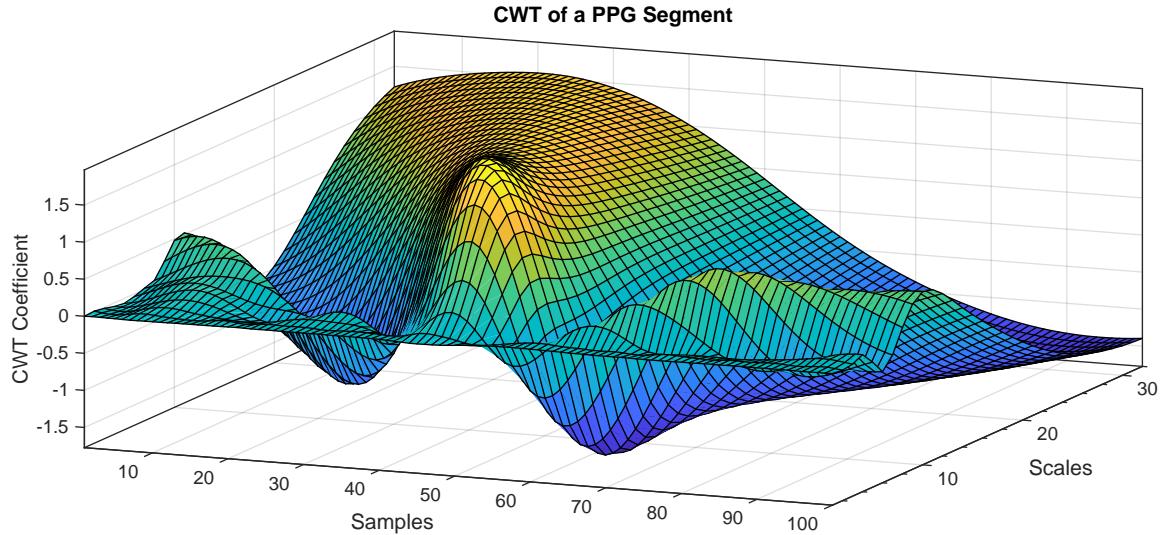


Figure 4.6: CWT transform of PPG segment

al. showed that 1DMRLBP had the following useful characteristics [62],

- Fast and computationally inexpensive to extract.
- Tolerant to shifting, scaling and misalignment noises.
- Provides more descriptive features by considering multiple resolutions.
- Considers temporal variations among heart beats.
- Accounts for quantization error.
- Online features, in the sense that, it does not require future observations.

Louis et al. achieved a remarkable 7.89 % EER on single session ECG database of 1012 subjects. In addition, when used in continuous authenticating, it achieved an astonishing 0.39 % False Rejection Rate and a 1.57% False Acceptance Rate [62]. Because ECG and PPG were one dimensional cardiac related signals and inspired by the success of 1DMRLBP on ECG biometrics, it was imperative for us to try 1DMRLBP for PPG too. In addition, 1DMRLBP offers all ideal properties that are crucial in any functional biometrics system. Compared to ECG, PPG is easily recorded and PPG sensors are already available in many wearable devices. Since 1DMRLBP was developed with a focus on wearable devices applications, it motivated us to try 1DMRLBP for PPG.

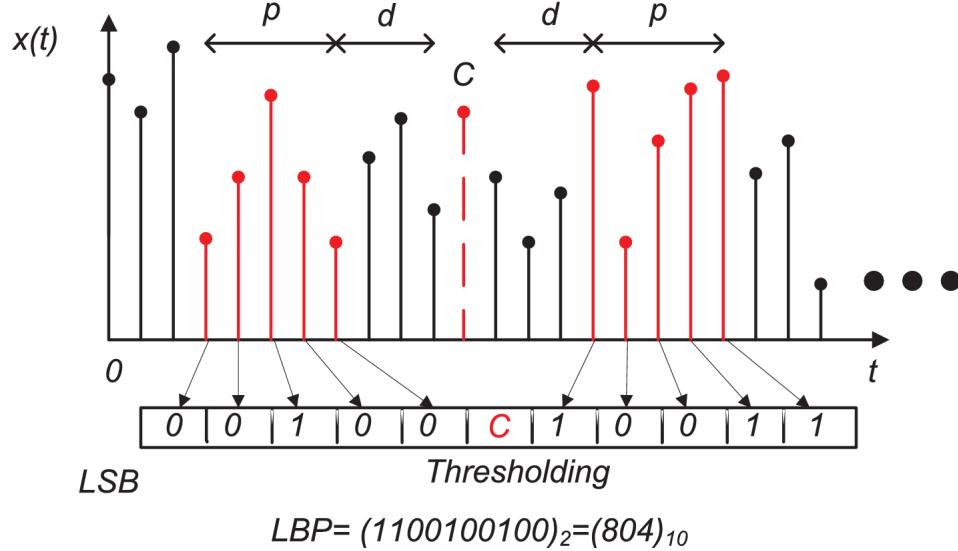


Figure 4.7: Example of 1DMRLBP with $d = 4$ and $p = 5$ [62]

4.2.2.2 1DMRLBP feature extraction

1DMRLBP was modified for 1D-signals and in particular for biometrics from common definition of LBP in computer vision and given as follows [62],

$$BP(x(t)) = \sum_{i=0}^p sign(x(t + i - p - d + 1) - x(t)).2^i + sign(x(t + i + d) - x(t)).2^{i+p} \quad (4.3)$$

$$\text{where, } sign(x) = \begin{cases} 1, & \text{if } x + \epsilon \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (4.4)$$

Here, $x(t)$ is a center time-sample, d is how far from center sample, Binary Patterns (BP) starts to get extracted, p is how many points to consider on each side of $x(t)$ for BP extraction. Fig. 4.7 shows an example of how 1DMRLBP is calculated. Different values of p and d enable the researchers to capture more signal characteristics at different resolutions for the same center sample $x(t)$. The value for BP is zero when Eq. 4.3 requires information that is out of bound for signal $x(t)$. That is for signal with length k , it becomes out of boundaries at $t + i + d > k$ and $t + i < p + d$. Parameter ϵ was introduced in Eq. 4.4 to account for quantization error. Scale in-variance is also achieved by Eq. 4.4. It was also argued that adding this leeway reduced influence of ripples in a heartbeat [62].

4.2.2.3 Accounting temporal variations through 1DMRLBP

Histogram operation was introduced to consider temporal variations among heartbeats, to eliminate spatial information, and to reduce noises due to shifting, misalignment, and segmentation. Normalized histogram

was calculated over respective overlapping/non-overlapping windows for BP values of past n heart beats.

Whole procedure is as follows [62],

1. For all n heart beats, calculate BP value for every time sample $x(t)$ according to eq. 4.3.
2. Calculate histogram of BP values over window of length w for all n beats.
3. Merge histogram from all n beats for particular window and normalize.
4. Concatenate normalized histogram with normalized histogram feature vector from previous window.
5. Shift window by Δ .
6. Repeat from step 2, until completion.

This formulation of capturing temporal variations results in one feature vector only, irrespective of value of n because of normalized histogram. Considering BP values over n heart beats allow for more descriptive features than one heartbeat. Discriminatory power of 1DMRLBP was explained to come from two facts. First, BP values for $x(t)$ encodes surrounding micro-texture patterns such as edges, valley, peaks, and inflection points etc. at multiple resolution. Secondly, calculating the frequency of such patterns through a histogram over multiple heartbeats accentuates subject specific patterns [62].

4.2.2.4 Example calculation for 1DMRLBP

In this work, PPG segment of 1s is considered with sampling frequency of 100 Hz. Hence, the length of each segment is $k = 100$ samples. Let's say $p = (4, 4)$, $d = (10, 10)$, $w = (50, 40)$, $\Delta = (10, 20)$. Since $p = 4$, 4 samples are considered for BP at distance of $d = 10$ for each $x(t)$. A total of $2^{2p} = 256$ binary patterns can be encoded with decimal values ranging from 0 to 255. Hence, in a histogram there are total 256 bins. For each window of length $w = 50$ with shift $\Delta = 10$, histogram is calculated. Total $l = 1 + \lfloor \frac{k-w}{\Delta} \rfloor = 6$, windows are formed. Histogram from each window is concatenated to generate 1536 features. Similarly for $w = 40$ and $\Delta = 20$, a total of 1024 features are generated. Final feature vector is formed at a length of 2560 by concatenating features from both. It is expected that most of the features would be zero giving sparse feature vector.

4.2.3 Auto-Correlation based features

Auto-Correlation (AC) based systems were state of the art for ECG based system for long time [63][64]. A benefit of using AC is that, it measures similarity in a signal by calculating sum of product (dot product again !) between the signal itself and its time-shifted version, as given in Eq. 4.5. AC reveals a presence of ‘periodicity’ or ‘quasi-periodicity’ effectively. It also embeds information such as the length of periods and distances between them without requiring to detect fiducial points of a PPG signal. Since the focus is on embedding information on periodicity, lengths, distances between different fiducial points etc., a PPG segment length more than one second is required to extract AC features. Further, systolic peaks detection

and alignment of PPG segments respective to them, is not required, as AC extracts information by shifting the signal by different lags m . Therefore, following ECG based systems, PPG is also blindly segmented into overlapping windows of predefined length. Then, normalized auto-correlation (AC) of each window was calculated as follows:

$$\hat{R}_{xx}[m] = \frac{\sum_{i=0}^{N-|m|-1} x[i] x[i+m]}{\hat{R}_{xx}[0]} \quad (4.5)$$

where, $x[i]$ represents windowed PPG segment, $x[i + m]$ shifted PPG segment with time lag of $m = 0, 1, 2, \dots, M - 1$, N is length of PPG window and $M \ll N$. AC features are particularly important for PPG, as it suffers respiration induced variations. From Fig. 4.4, it can be seen that PPG segments are not completely overlapped for the same subject and they have small time variations among them. AC along with feature reduction techniques described in next section can provide features that are immune to these variations.

4.3 Feature Reduction

4.3.1 Motivation for feature reduction

In most of the biometrics applications, dimensionality of feature vector is generally very high. In such cases, system suffers from the curse of dimensionality. With high dimension feature vector, an exponentially larger number of training samples are required compared to lower dimension feature vector [65]. In biometrics, the primary task is to distinguish between different subjects, thus to identify features that take different values among subjects. However, high dimensions feature vectors carry a redundant amount of information that is similar across different subjects. In such cases, it is believed that feature vectors lie on lower dimensional subspace in high dimensional feature space. Projecting feature vectors on this lower dimensional feature space could greatly reduce computations, storage, and the number of training points required to learn classifier. These benefits are of huge importance for a practical biometrics system where it could lead to shorter training time, authentication time, less storage, and processing requirements. Being a physiological signal, PPG changes with time and finding subspace could lead to a better performance. In the interest of biometrics, we would like to find a lower dimension subspace such that, different subjects are ‘well separated’ to each other in that subspace. We revert to variants of Linear Discriminant Analysis (LDA) to achieve the purpose of finding such a subspace.

4.3.2 Linear Discriminant Analysis (LDA)

LDA is a supervised technique that finds linear subspace such that it increases the inter-subject variability and reduces the intra-subject variability in that subspace. In other words, LDA finds the direction of a subspace where centers of each class are as far apart as possible while keeping each class’s variance as small as possible. Formally, given a training set $Z = \{Z_j\}_{j=1}^K$ containing K classes each having $Z_k = \{\mathbf{z}_{ki}\}_{i=1}^{N_k}$ where N_k is the number of feature vectors for each class k , LDA finds projection weight \tilde{W} that maximizes the

fisher criterion function $\mathbb{J}(W)$:

$$\tilde{W} = \arg \max_W \mathbb{J}(W) = \arg \max_W \left(\frac{|W^T S_b W|}{|W^T S_w W|} \right) \quad (4.6)$$

here S_b and S_w are the between class and within class scatter matrix respectively defined as:

$$S_b = \sum_{k=1}^K N_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T \quad (4.7)$$

$$S_w = \sum_{k=1}^K \sum_{i=1}^{N_k} (\mathbf{z}_{ki} - \boldsymbol{\mu}_k)(\mathbf{z}_{ki} - \boldsymbol{\mu}_k)^T \quad (4.8)$$

where $\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{z}_{ki}$ is the mean of class Z_k and $\boldsymbol{\mu}$ is the overall mean vector. It can be observed that the objective function $\mathbb{J}(W)$ is invariant to scaling of W . Furthermore, we are only interested in finding the direction of W . Thus, optimization can be rewritten as,

$$\max(W^T S_b W) \quad s.t. \quad W^T S_w W = 1 \quad (4.9)$$

using Lagrange multiplier it can be written as following: (4.10)

$$\mathbf{L}(W, \lambda) = W^T S_b W - \lambda(W^T S_w W - 1)$$

taking differentiation w.r.t W and setting it to zero,

$$\frac{\partial \mathbf{L}(W, \lambda)}{\partial W} = 0 \implies S_b W = \lambda S_w W \quad (4.11)$$

Eq. 4.11 is a generalized eigenvalue problem. If S_w is not singular then, LDA finds \tilde{W} as $m = K - 1$ most significant eigenvectors of $(S_w)^{-1}(S_b)$ that corresponds to first m largest eigenvalues. Here, rank of S_b is at most $K - 1$, while S_w is full rank. Therefore, $(S_w)^{-1}(S_b)$ has at most $K - 1$ non-zero eigenvalues and LDA is limited only to find subspace at most of dimensions $K - 1$. After obtaining weight \tilde{W} , all the input training PPG segments are subjected to linear projection of $\tilde{\mathbf{z}} = W^T \mathbf{z}$. Projected training vectors $\tilde{\mathbf{z}}$ together with weight \tilde{W} are saved in gallery for template matching or classifier training.

4.3.3 Small Sample Size Problem

In practice, due to the high dimensionality of feature vectors and the low number of training samples, S_w is almost always singular. For example, each PPG CWT feature vector is of size 3200 and S_w is of size $3200 \times 3200 = 10$ M. This means for non-degenerate S_w , atleast 10 M training samples are needed. Also, such a big matrix is computationally expensive to handle. This problem is called Small Sample Size (SSS) problem and LDA becomes infeasible in such cases.

Problem of SSS was mainly studied in the context of face recognition where only a few number of training samples are available with high dimension feature vectors. Solutions are available in literature which were mainly developed for face recognition but are equally applicable to our PPG biometrics system too. One of the popular solution is to apply PCA as an intermediate step before LDA to reduce dimensionality [66].

However, a major drawback of this solutions is that, PCA might discard useful discriminative information. Another solution is to add a small amount of noise in the covariance matrix such that it becomes non-singular [67].

4.3.4 Direct LDA

In this work, Direct-LDA was used to overcome problem of small sample size [68]. In traditional LDA if S_w is singular then to proceed, null space or eigen vectors corresponding to zero eigen values are discarded. However, it was argued that, the null space of S_b contained no useful information and could be discarded rather than discarding the null space of S_w which contained the most useful information. Intuition, behind this was that for a projection direction \mathbf{a} if $S_w\mathbf{a} = 0$ and $S_b\mathbf{a} \neq 0$ then Fisher's criterion $\mathbf{a}^T S_b \mathbf{a} / \mathbf{a}^T S_w \mathbf{a}$ was maximized in that direction and prefect classification could be achieved. If $\mathbf{a} \in$ null space of S_b then $\mathbf{a}^T S_b \mathbf{a} = 0$, hence null space of S_b was considered not to carry any useful info.

Direct-LDA (DLDA) starts with the same objective as traditional LDA, to find W such that it simultaneously diagonalizes both S_w and S_b while maximizing Fisher's criterion that is, $W^T S_b W = \Lambda$ and $W^T S_w W = I$. To achieve this objective, first S_b is diagnoлизed and the null space of S_b is discarded. Then S_w is diagonalized using subspace found earlier, such that it retains eigenvectors 'close to zero', while maximizing Fisher's criterion. This algorithm is called unified LDA or direct-LDA as it does not involve any intermediate steps such as PCA or adding perturbations. Computations for eigen vectors of scatter matrices ($N \times N$) are greatly reduced by observation presented in EigenFace approach [69]. For example between class scatter matrix can be written as following,

$$S_b = \sum_{k=1}^K N_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^T = \Phi_b \Phi_b^T (N \times N) \quad (4.12)$$

$$\Phi_b = [\sqrt{N_1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}), \sqrt{N_2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}), \sqrt{N_3}(\boldsymbol{\mu}_3 - \boldsymbol{\mu}), \dots] (N \times K) \quad (4.13)$$

where N is dimension of feature vector and K is number of classes. Calculating eigen vectors of, $N \times N$, $\Phi_b \Phi_b^T$, is computationally expensive. Instead, first eigen vectors of $\Phi_b^T \Phi_b \mathbf{v}_i = \lambda_i \mathbf{v}_i$ are calculated. Multiply both, sides by Φ_b resulting in, $\Phi_b \Phi_b^T \Phi_b \mathbf{v}_i = \lambda_i \Phi_b \mathbf{v}_i$. Replace $S_b = \Phi_b \Phi_b^T$ and $\mathbf{u}_i = \Phi_b \mathbf{v}_i$ resulting in, $S_b \mathbf{u}_i = \lambda_i \mathbf{u}_i$. This suggests that, eigen values of S_b are same as $\Phi_b^T \Phi_b$ and eigen vectors are linear combination of \mathbf{v}_i . Thus eigen vectors of S_b can be efficiently calculated from $K \times K$, $\Phi_b^T \Phi_b$ matrix.

4.3.5 Motivation for Non-Linear Method

DLDA is a linear method, however it is possible that the distribution of PPG signal under different physical and mental states is non-linear. The exact distribution is difficult to estimate. Considering this hypothesis is true, linear DLDA will struggle to provide a reliable solution. To tackle this issue, non-linear or kernel versions of LDA were also explored in this work. The central idea behind this is to use 'kernel trick' to implicitly map input space to implicit high dimensional feature space, where it is believed that, distribution

is linear and simplified. To do so, all equations are reformulated in terms of inner products and replaced with values from kernel matrix. Kernel matrix can easily be computed from input feature space vector. However, in high dimensional space, SSS problem becomes severe. Solutions like Kernel PCA (KPCA) + LDA and Generalized Discriminant Analysis (GDA) were proposed in the literature [70][71]. In KPCA + LDA, LDA is performed on the subspace found by KPCA while GDA is the same as LDA except computations are done in high dimensions. Being exactly similar to their linear versions, these methods suffer from the same limitations as their linear versions.

4.3.6 Regularized Kernel Discriminant Analysis (R-KDA)

In this thesis, Regularized Kernel Discriminant Analysis (R-KDA) proposed by Lu et al. was used for non-linear LDA [72][73]. According to Lu et al., in extreme SSS scenarios, DLDA suffered from high variance in the estimation of null space of S_w [74][72]. In DLDA, to avoid division by zero eigen values from S_w either only close to zero eigenvalues or smallest eigenvalues boosted to some predefined value were used. In doing so, null space of S_w which contained the most discriminatory information could not be utilized completely. To avoid these problems in R-KDA, Lu et al. considered the following modified fisher's criterion,

$$\tilde{W} = \arg \max_W \mathbb{J}(W) = \arg \max_W \left(\frac{|W^T S_b W|}{|\eta(W^T S_b W) + (W^T S_w W)|} \right) \quad (4.14)$$

where η is a regularization parameter between [0,1]. It was proved that modified criterion was equivalent to Eq. 4.6 and produced the same \tilde{W} as the original criterion. Additionally, η helped reduce variance in estimation of eigenvalues and setting $\eta > 0$ avoided division by zero in finding subspace. This way information from both inside and outside of the null space of S_w could be extracted while avoiding variance and division by zero. For the eigen analysis in implicit high dimensional space, the same strategy as DLDA was followed with modified criterion while exploiting the kernel trick, various properties of the kernel matrices, and by utilizing Eq. 4.12. In this work, Radial Basis Function (RBF) kernel $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}\right)$ was used. For R-KDA η and σ are hyper parameters tuned during experiments.

4.3.7 Other feature reduction techniques

During this work, few other feature reduction methods were also considered, such as Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), Locally Linear Embedding (LLE) and Isomap. PCA and kernel PCA are both unsupervised techniques which try to minimize the reconstruction error of data points by retaining direction of maximum variance. The PCA approach is oblivious to class labels and projects data into directions of maximum variance which does not guarantee good classification. LLE and Isomap are manifold learning techniques which also do not consider class labels. Our experiments with these techniques resulted in poor performance, which is not surprising given the objective of these techniques. Supervised variants of these methods are also available in literature. However, we find that, LDA based methods are best suited for our application, as its objective function is most practical. Unsupervised techniques can also be useful in situations where one wants to build a biometrics system using only data

from one subject. This concept is also referred to as one-class learning. In the context of biometrics, one class learning builds a decision boundary around data from genuine subjects, and anything outside it is considered as imposter. This direction was not pursued in this thesis. For brevity's sake, further discussion on unsupervised learning is avoided herein and experimental results are not included for comparison, as it would be poor comparison to make.

4.4 Classification

During the enrollment phase, after passing training data from pre-processing, feature extraction and feature reduction blocks, a classifier is trained for each subject. In this thesis, focus is on the verification mode of biometrics that is given a query signal, the system should either accept or reject it. Experiments with many different classifiers were carried out. Classifiers explored during our work can be roughly categorized in two ways: distance based classifiers and decision boundary or rule based classifiers. As part of the first category of classifiers, Nearest Neighbor, Correlation Distance, Log Likelihood Ratio Test (LLRT) [75], and Nearest Neighbor Data Descriptor (NNDD) [76] were examined. While in other category, Support Vector Machine (SVM) and Decision Trees were examined. Results showed that, distance based classifier performed poorly compared to SVM and Decision Trees. Poor performance can be attributed to the fact that, any outlier in training data could greatly influence the outcome of a classifier. Other drawbacks are that it requires storage of training samples which might not be practical and and that all these classifiers are, furthermore, linear. Distance based classifiers can simply be seen as ‘template matching’, which does not optimize, exploit or transform the structure of training data for better classification performance. Although, there exist solutions in literature to address these shortcomings, exploring such solutions is out of scope of this thesis. Hence, discussion on distance based classifiers and comparison of their results is further avoided. The following sections provide a review on SVM, decision trees and difficulties involved in training these classifiers.

4.4.1 Support Vector Machine (SVM)

Support Vector Machine is a theoretically motivated powerful classifier which focuses on minimizing generalization error by finding decision boundary that maximizes the margin between two classes [77][78]. Innately, SVM is a binary classifier that given class labels $y_i \in \{-1, 1\}$, training vectors $\mathbf{x}_i \in \mathbb{R}^d$ for $i = 1 \dots N$, it finds hyperplane \mathbf{w} that maximizes the margin by solving the following constrained optimization,

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (4.15)$$

$$\text{such that, } y_i(\mathbf{w}^T \mathbf{x}_i - b) \geq 1 - \xi_i \text{ for } i = 1, 2, \dots, N.$$

Here, $\xi_i \geq 0$ are called slack variables which allow for some data points to be inside the margin ($0 \leq \xi_i \leq 1$) or allow to be misclassified ($\xi_i > 1$). Thus $\sum_i \xi_i$ provides an upper bound on training error. Parameter $C > 0$ controls the trade-off between minimizing the training error and maximizing the margin [78]. In

this light, C can be seen as the inverse of regularization parameter and Eq. 4.15 can be seen as minimizing for hinge loss [79]. Eq. 4.15 can best be optimized in its dual form, where it can be observed that, dual formulation only depends on input data through dot products. Hence, ‘kernel trick’ can be employed to generate non-linear decision boundary. After obtaining a solution of dual with Lagrange multipliers a_i and bias b , decisions on any test sample are made by sign of following,

$$y(\mathbf{x}) = \sum_{i=1}^N a_i y_i \mathbf{x}_i \cdot \mathbf{x} + b \quad \text{and for non-linear case} \quad (4.16)$$

$$y(\mathbf{x}) = \sum_{i=1}^N a_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \quad \text{where } k(\mathbf{x}_i, \mathbf{x}) \text{ is computed from chosen kernel.} \quad (4.17)$$

From Karush Kuhn Tucker (KKT) conditions, it is also revealed that for some data points $a_i=0$. From Eq. 4.16, such data points are not useful in making decisions. Hence, after training these data points can be discarded [78]. This property is of particular importance for practical biometrics system where storage may be a limitation. For the rest of the data points for which $a_i > 0$ are called support vectors. In this work, experiments with both linear and non-linear versions of SVM were carried out. For non-linear SVM, RBF kernel, $k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}\right)$ was used where σ is the width of RBF kernel tuned during training. Experimental results showed that, RBF SVM performed better than linear SVM. Hence, for brevity, only results from RBF kernel are included and compared in this thesis. For training SVM, One Vs All (OVR) approach was used where data from genuine subject formed the positive class and rest of the subjects formed the negative class. Interestingly, looking at Eq. 4.17 for RBF kernel, interpretation of classification can be made as the accumulation of scores by RBF bumps centered at support vectors on test sample [80].

4.4.2 Addressing class-imbalance problem in training SVM

In training SVM, one-vs-rest (OVR) approach was employed. A major problem with OVR is that the training data is highly imbalanced. For example, if we had 40 negative classes with 100 points and against one positive class with 100 points then the total number of negative points (majority class) (4000) are much greater than positive points (minority class)(100). If a decision boundary is drawn such that it classified all points as the negative class during training then also, training accuracy of 97.56% would be achieved. In small sample size scenario with large number of classes, this problem becomes severe. In such situations, accuracy turns out to be a flawed evaluation metric as it gives equal importance to the classification of majority and minority samples. A simple way to address this problem is to use other evaluation metrics such as Precision, Recall, F-Measure, Area Under the Curve (AUC) or ROC (Receiver Operating Curve) [81]. However useful these metrics are, they do not reflect in training phase of a classifier. Hence, it is possible that, the classifier would still be biased towards the majority class.

Learning from imbalanced data is an active research area of its own. To overcome this problem of class imbalance during the training phase, several techniques based on sampling methods, cost sensitive framework and one-class learning based approaches have been proposed in the literature [81]. In sampling based

approaches, either the minority class is over-sampled or the majority class is under-sampled or alternatively, both are done to some extent such that training data is not imbalanced for training classifier. Under sampling refers to selecting only subset of majority samples for training data while over sampling refers to either synthetically generating majority samples or just using copies of them. The problem with this approach is that, one needs to identify good representative majority samples such that classifiers would generalize well on unseen data. Additionally, to over sample the minority class in a meaningful way without being susceptible to outliers (variance problem), one should have a good estimate of distribution of minority class. In the context of biometrics, one would want to use as many negative samples as possible to train classifier to reduce false positive error, under the limitation of small amount of training samples from genuine subject. Hence, the sampling approach is not preferred.

During our experiments, we tried the popular Synthetic Minority Oversampling Technique (SMOTE) which generates more minority samples by randomly selecting a point on the line, connecting training data to its k -nearest neighbors. The amount of over sampling can be controlled by parameter k [82]. Optional but recommended is to under sample the majority class simultaneously. However, initial results showed poor performance, which could be attributed to reasons mentioned earlier. Hence, this line of direction was abandoned. Many other variants of SMOTE and other sampling techniques are also available [81]. But in interest of biometrics application, these options were not explored further.

Another approach to solve class-imbalance is of one-class learning. One-class SVMs were developed mainly for outlier or novelty detection where a large number of negative samples were available before hand and the task was to detect any future positive sample or an outlier. SVM was then trained only on negative class. Two of such popular techniques are ν -SVM and Support Vector Data Descriptor (SVDD) [83][84]. ν -SVM finds a hyper-plane such that training data is maximally separated from the origin while SVDD minimizes volume of hyper-sphere enclosing training data. During our experiments, we tried training ν -SVM and SVDD with data from only genuine subject, but resulted in poor performance. The reason for that is, in the absence of training imposter samples, it produces closed decision boundaries and results in less generalization. Another reason is that careful initialization of parameters is required. Author of SVDD also argued that if few example imposter samples were included during training then better results could be achieved [84]. However, that begs the same question as earlier that is to find good representative imposter samples. Hence considering all these, one-class learning approach was also abandoned.

To solve the class imbalance problem, a cost sensitive approach was used in this thesis. For SVM this is achieved by assigning different costs for mis-classifying majority and minority class samples. To accommodate this, the second term in Eq. 4.15 is modified to the following [80]

$$C \sum_i^N \xi_i = C_+ \sum_{i \in I_+} \xi_i + C_- \sum_{i \in I_-} \xi_i \quad (4.18)$$

here C_+ and C_- are penalties for mis-classifying positive and negative samples respectively. I_+ and I_- are set of positive and negative samples. To give equal overall costs to each class, C_+ and C_- are set according

to following,

$$C_+ n_+ = C_- n_- \quad (4.19)$$

$$\frac{C_+}{C_-} = \frac{n_-}{n_+} \quad (4.20)$$

where n_+ and n_- are total number of positive and negative samples. This way, we could avoid problems associated with other techniques while utilizing all available imposter training samples. Additionally, instead of accuracy as an evaluation metric, FAR, FRR and ROC are used to provide robust assessments under class imbalance.

4.4.3 Random Forest based Classification

Simply put, Random Forest is an ensemble of randomly trained decision trees. A decision tree is a hierarchical piece-wise model which breaks down complex problems into simpler ones at each stage [85]. However, a single decision tree has a high variance problem that is given a slight change in training data, it produces completely different results. The reason is, error made at the top nodes of a tree propagate all the way down [79]. To alleviate this problem, Random Forest bets on the ‘wisdom of crowd’ and ‘randomness’ to generate powerful classification model. Randomness is injected in two ways while training [85],

1. Each tree is trained on bootstrapped samples generated by sampling training data uniformly at random with a replacement.
2. For selecting splitting feature for each node, it only considers a subset of features selected at random from all available features.

Randomness makes sure that each tree is trained on different training data and all trees are *de-correlated*. Decision for test samples are made by averaging decision from all trees. Another great aspect is even with linear weak learners, it produces global non-linear boundaries, due to the fact that multiple simple linear split nodes are arranged in hierarchy. In addition, it was also shown that just like SVM, random forest also maximized the margin under certain conditions [85] [86]. Random forest also optimizes for information gain while training. This way it is also doing feature selection.

In this work, we used the recommended value of $\lfloor \sqrt{d} \rfloor$ where d is the dimensionality of feature vector, as a size of subset while randomly selecting feature [87][79]. It was also found that, 51 trees were enough for an ensemble. For training, standard CART algorithm with the gini index as in Eq. 4.21 was used as impurity/entropy metric [88].

$$GDI = 1 - \sum_i p_i^2 \quad (4.21)$$

where i stands for number of classes, which is two in our case and p_i is observed fraction of samples for class i . To address class imbalance issue, mis-classification costs with the same proportion as shown in Eq. 4.19 were used. It has an effect of oversampling minority samples while generating bootstrapped samples [79]. This way, it was made sure that no tree was just trained on majority samples by accident. Sampling here

does not suffer problems mentioned earlier as all imposter samples are used during training. In fact, by nature, bootstrapping procedure performs sampling.

Chapter 5

Empirical Results

This chapter begins with the problem of parameter selection, trade-offs, computational complexity and multiple choices that one faces during each module design while devising and tuning whole system for best performance. Discussions related to various errors and limitations of each module is presented. The effects of these limitations on performance is analyzed qualitatively. Subsequently, results are presented for each database and its configurations after finalizing on various parameters. Results are analyzed and supported with insights.

5.1 Difficulties in engineering whole pipeline

As discussed in previous chapter, our biometrics system model is a typical biometrics or classification system consisting of signal recording, pre-processing, feature extraction, feature processing, classification, and finally decision. However, tuning whole pipeline from signal recording to decision for best performance is an extremely difficult task because of various errors and trade-offs.

In theory, one only focuses on tuning one module of a system while keeping others fixed. For instance, tuning classifier while keeping everything else fixed. However, it is not hard to imagine that if some other module's parameters are changed then, previously tuned modules may not provide optimal performance. Hence, when looking at a system in its entirety, many trade-offs are faced. Finding the best parameters or tuning all parameters for whole system is an enormous 'combinatorial task'. Although, one can rely on heuristics, recommendations, literature, constraints etc., to reduce the search space, then even the various trade-offs do not vanish.

Apart from that, each module/technique injects its own errors and has its own limitations. For example, peak detection algorithm depends on parameters set empirically to give the best results. However, in adverse conditions, it is not hard to imagine that these parameters values may provide false peaks. In addition, signal segmentation depends on peak detection. Segmentation under heart rate variability and respiration

induced noise would introduce misalignment and segmentation error. Outlier segments removal also depend on thresholds to remove segments that are severely deviated from the median segment. Errors introduced previously can also affect the operation of the outlier removal. These errors further propagate to feature extraction, feature reduction and classifier stage.

At feature reduction stage, DLDA is known to have problems of high variance in estimating eigenvalues of S_w in extreme Small Sample Size (SSS) scenario e.g. training with only 2 – 6 training samples each having high dimensions. Because of errors in the pre-processing stage, if training set of PPG segments falls into the category of extreme SSS then, DLDA would also inject errors into the system. Fortunately, most of the time, we did not have extreme SSS scenarios during training as each subject had more than 60 training segments. Hence, we expected that our system would not suffer this limitation. It was further verified by comparisons of results with R-KDA presented in later sections.

On the other hand, R-KDA performance heavily depends on the regularization parameter η and kernel parameter σ . Finding the best values of these parameters is computationally extremely expensive. For example, for the single session configuration of BioSec.Lab PPG database, there are 84 subjects and training set consisted of approximately total of $M=8100$ segments from all subjects. In this case, R-KDA requires computation of kernel matrix $k(\mathbf{z}_i, \mathbf{z}_j) = \exp\left(-\frac{\|\mathbf{z}_i - \mathbf{z}_j\|^2}{\sigma^2}\right)$ of size $M \times M$ that is $8100 \times 8100 \approx 65M$ values. One could use symmetricity of kernel matrix and half the computations. But even then, asymptotically it is equally computationally expensive. Finding optimal subspace requires manipulation of this huge kernel matrix. If that is not enough, kernel matrix computation depends on kernel width parameter σ . Value of σ spans across a huge range from $[0, \infty)$. For $\sigma \rightarrow 0$ then, $k(\mathbf{z}_i, \mathbf{z}_j) \rightarrow 0$ for $i \neq j$, and mapped samples tend to be orthogonal to each other. When $\sigma \rightarrow \infty$, then $k(\mathbf{z}_i, \mathbf{z}_j) \rightarrow 1$ and all points collapse to one single point. Hence, a value of σ has a great effect on the separability of classes and its value should not be too large or too small [89]. Finding optimal value of σ on such a huge range with large kernel matrix is computationally very expensive. Another difficulty here is that as training data changes, the optimal value of σ also changes. For example, the best value of σ found for single session configuration is not necessarily the best value for the across exercise configuration as training data is different. Additionally, along with σ , regularization parameter η also requires tuning. Value of η spans across $[0, 1]$. Simply considering a combination of 3 values of η and 10 values of σ would require to compute kernel matrix 30 times, for single session that would be $30 \times 65 \approx 1.9B$ kernel matrix values. Also, feature vector \mathbf{z}_i itself has high dimensions, e.g. for CWT features it is of dimensions 3200. Hence, the actual number of primitive operations in fact is even larger and that would be for the single session configuration only while keeping the rest of the systems fixed. Considering that the rest of the system also needs to be tuned along with R-KDA, it becomes unfeasible.

Both DLDA and R-KDA also depend on hyper-parameters to adjust eigenvalues during the process. These parameters were left untouched and default values were used. In this case, DLDA is completely parameter-free and much faster computationally than R-KDA. Hence, during experiments, DLDA was first used to select best parameters for the rest of the system. Then, experiments with R-KDA were carried out only to improve upon the performance of DLDA. However, under sub-optimal values of η and σ , R-KDA is not guaranteed to improve performance. To reduce the search space for σ , grid search over range of $[10^{-3}, 10^{17}]$ with $\eta = 0.001$ was performed over a small subset of data. It was found that σ over range

$[10^5, 10^7]$ performed well. Hence, in later experiments with all database configurations, only three values of $\sigma = 10^5, 10^6, 10^7$ with a combination of three values of $\eta = 10^{-3}, 10^{-4}, 10^{-5}$ were used to find the best set of values.

Given these difficulties in finding the best values of parameters for kernel methods, it is not surprising that in literature many ways to address this problem are also available. These methods include techniques that directly optimize kernel matrix or kernel function using some criterion. For example, in [90] Semi Definite Programming (SDP) is used to optimize kernel matrix by maximizing margin in kernel feature space. Although most of these methods eliminate the need to search for best kernel parameters over vast ranges, these methods themselves depend on many optimization parameters, their initial values, constraints, etc. Hence, the need for selecting the best kernel parameters values is probably simplified at the cost of complexity but the need for parameter selection is not completely eliminated. Therefore, in the interest of only assessing the biometrics capabilities of PPG, this direction was not further investigated during this work. In real life, one can fine tune using these methods.

For SVM at the classifier stage, the one-vs-all approach was used. SVM performance depends on penalty parameter (C) and kernel width (σ). Generally, a cross validation approach is used to find the best values for these parameters. If training with the single session configuration of BioSec.Lab PPG database, this would require to train 84 SVM classifiers, one for each subject. For all configurations and all databases, training one-vs-all classifier using cross-validation increasingly becomes computationally expensive. Additionally, parameters C and σ found using cross-validation here depend on input data fed into SVM during training. Hence, if data changes then these parameters' values found earlier does not serve as optimal values. Now, given one also has to find many parameters values in earlier stages, using cross-validation to find SVM parameters every time increasingly becomes unfeasible. Therefore, in this work, we did not use cross-validation to find values of C and σ . We relied on MATLAB's in-built heuristics procedure to determine a value of σ , and penalty parameters were set according to cost matrix described in previous chapter. We also experimented with Bayesian optimization to reduce cross-validation loss to find the best values for σ and C for few trials. However, it did not make much of a difference in results while being computationally very expensive to carry out. On the other side, MATLAB's heuristic procedure uses sampling to determine σ , hence carries randomness in estimation of σ each time making parameter selection difficult. We controlled the random number generator for the reproducibility of results.

Additionally, for SVM, a particular value of C and σ decide the training error and generalization. Kernel width σ determines the flexibility of decision boundary in kernel space while penalty parameter (C) controls the margin and orientation of decision boundary [80]. In our cost-sensitive framework, penalty parameter C_+ on positive class is set much higher than negative class to adjust for class imbalance. However, in this case positive samples closer to boundary would affect the orientation and margin of decision boundary at large. Therefore, it is not difficult to guess that during training large value of penalty parameter C_+ along with σ would possibly over-fit.

For, SVM objective optimization, Sequential Minimal Optimization (SMO) solver of MATLAB was used. SMO solver is also not perfect and has its own limitations. For example, the smallest working set selection and numerical issues. During optimization, many comparisons are made for Lagrange multiplier a_i . Because

of finite difference approximation errors, at each iteration, these comparisons could lead optimization to completely different paths, thereby producing sub-optimal results [91]. Because of all these errors and approximations, parameters selection was further made difficult.

Similarly, for Random Forest, parameters such as the number of trees in the ensemble, max number of splits, depth of tree, number of features to select at random at each split etc. need to be tuned for the best performance. Here also, the one-vs-all setting has to be used and general mechanism of using cross-validation is computationally expensive. For simplicity, we relied on default values of MATLAB for all these parameters. The number of trees in the ensemble were set to 51 for all classifiers. It should be noted that, this number was set as an odd in order to avoid arbitrary selection in case of a tie at the end for the binary classification. Moreover, the number of trees value in the ensemble was set empirically. Therefore, it is not hard to imagine that depending on the distribution of training data, for few one-vs-all trials, this number may over-fit while for few possibly under-fit, leading to sub-optimal results. However, fine tuning every one-vs-all classifier every time is again computationally very expensive. Furthermore, by nature, random forest outcomes contain statistical variations, making parameter tuning difficult.

Hence, considering all the factors mentioned above, accepting the limitations and difficulties in training and tuning whole system together, it is justified to consider results included here as only the *Empirical Results*. Given these difficulties and intricate implementation details, program code for generating results included in this chapter will be made public for anyone to experiment further and reproduce results. This would enable more transparency and propel more research on PPG Biometrics ¹.

5.2 Experiments with BioSec.Lab PPG Database

BioSec.Lab PPG database contains data recorded primarily in four different configurations i.e. single session, across exercise, short time-lapse and long time-lapse to evaluate permanence and the robustness of PPG biometrics. A detailed description of each configuration can be found in Chapter 3. Extensive experiments with each configuration were carried out with three feature extraction techniques that were 1. Continuous Wavelet Transform (CWT) 2. One Dimensional Local Binary Patterns (1DMRLBP) and 3. Auto-Correlation (AC). The following sections present results from each feature extraction technique for each configuration.

For every experiment, the system was trained using PPG data from the first 90 second of the training session. Testing was carried out with the first 20 segments selected from testing session. It was made sure that both the training and testing set were mutually exclusive. Prior to training and testing, PPG segments were generated by passing PPG signal through same pre-processing pipeline as described in Chapter 4. The final decision on test segments was made by majority voting.

For the single session, training and testing session were the same. That is, the first 90 second of three minute signal was used for training while the rest for testing. For the rest of the configurations, training and testing sessions were different. For example, in across exercise configuration, training was carried out using

¹https://github.com/umangyadav/PPG_Biometrics_BioSecLab

data recorded in relaxed condition, while testing was done using data recorded just after exercise. Similarly, for the short time-lapse, training session was the first session while testing session was the second session data recorded after atleast 30 minutes. This way, we could analyze the effect of time-difference and exercise on the performance of PPG biometrics system by comparing results with the single session. However, it should be noted that the number of subjects for all these configurations are not the same. Hence, directly comparing results with one another is not fair. Since EER is a normalized performance metric, for simplicity, we regard these results as an estimate of performance under different configurations. Results comparisons are justified in that sense.

For the long time-lapse configuration, few subjects had data available from a total of three sessions. Hence, for experimental convenience, it was divided into a total of four configurations. The first one being two sessions long time-lapse where data from 37 subjects were available from two sessions. Here, again the training session was the first session, while the testing session was the second session recorded after 36 days on average. Secondly, for 16 subjects, the third session recording was also available. Hence, for these 16 subjects, three more experiment configurations were formed. 1. Training on the first session, testing on the second session, 2. Training on the first session, testing on the third session, 3. Training on the second session, testing on the third session. These three new configurations have same number of subjects and can provide better insights on effect of time-lapse on PPG biometrics. Hence, in total seven different configurations were available for experiments.

For each feature extraction technique, the same system model and evaluation framework was utilized. As mentioned previously, DLDA was first used to fix the system parameters, then experiments with R-KDA were carried out only to improve upon results with DLDA. However, it is not guaranteed that R-KDA will always improve performance because of various factors discussed previously. Similarly, both classifiers SVM and Random Forest also have their limitations. Therefore, for the sake of only assessing capability of PPG as biometrics, only the best performing combination of feature reduction and classifier is used at the end to compare all results across different configurations and feature extraction techniques.

5.2.1 Results with Continuous Wavelet Transform (CWT)

CWT features primarily depend on choice of mother wavelet and a number of scales used to generate the time-frequency representation of a signal. Detailed discussion on the operation of CWT, effect of selection of scales and mother wavelet can be found in Chapter 4. Experiments with various combinations of mother wavelets and a number of scales using DLDA with both classifiers were carried out to select the best parameters. For example, we experimented with a number of chosen scales. They include following:

```
1 : 4   1 : 8   1 : 16   1 : 32   1 : 64   1 : 128   1 : 2 : 16   1 : 2 : 32   1 : 2 : 64   1 : 2 : 128,  
      1 : 4 : 16   1 : 4 : 32   1 : 4 : 64   1 : 4 : 128   10 : 4 : 136
```

Here, scales are represented in form of start:step:end. The choice of the number of scales determines the dimensionality of the feature vector. For example, for one second of the PPG segment with 100 Hz sampling

Table 5.1: Results with CWT as features. Columns represent feature reduction technique with classifier, Rows are different configurations from database, Cell values are performance in EER. For, R-KDA, results are from three different values of parameter σ . Best Results are highlighted in green.

		DLDA		R-KDA with $\eta = 0.0001$			
		Random Forest	RBF SVM	Random Forest	1.00E+07	1.00E+06	1.00E+05
Single Session	3.94%	4.76%	4.57%	3.57%	4.55%	3.57%	4.76%
Short Time Lapse	31.23%	29.15%	32.49%	32.57%	31.14%	30.91%	33.07%
Across Exercise	45.04%	47.80%	43.23%	40.69%	41.39%	44.80%	47.85%
Two Sessions Long Time Lapse	33.15%	38.71%	31.60%	35.22%	37.01%	38.64%	38.79%
Three Sessions Long Time Lapse	Train on 1st, Test on 2nd	38.53%	50.00%	38.19%	37.70%	35.94%	50.42%
Long Time Lapse	Train on 1st, Test on 3rd	30.16%	37.50%	34.69%	36.32%	35.24%	40.42%
	Train on 2nd, Test on 3rd	35.18%	31.25%	39.63%	35.54%	35.44%	33.75%

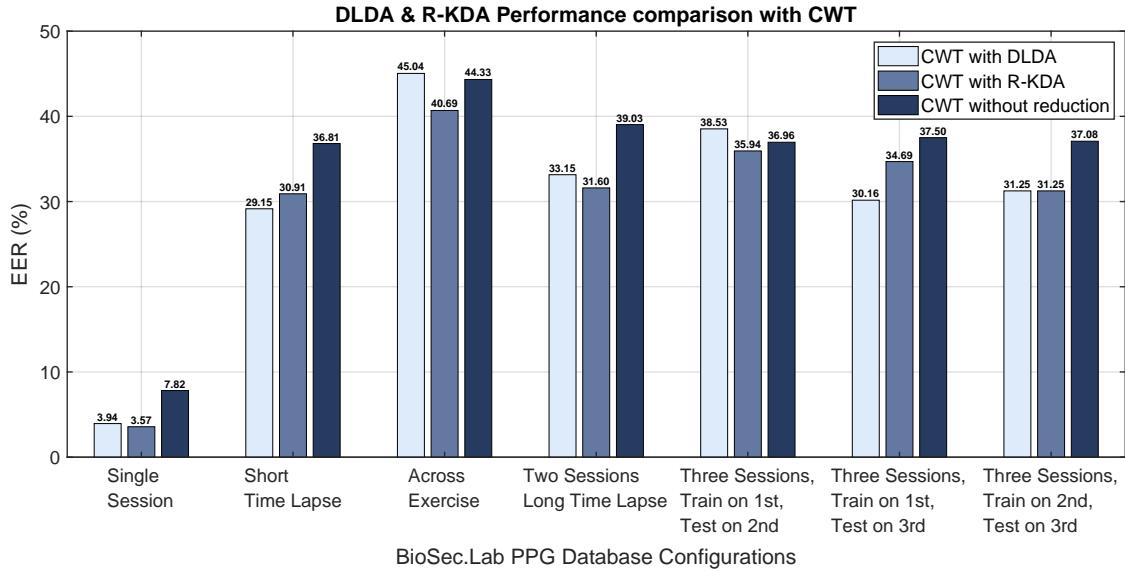


Figure 5.1: CWT performance comparison by taking best result each feature reduction technique

frequency, scales 1:4:128 would result in 3200 dimensions feature vector. It was found that, if a sufficient number of scales were chosen, then the actual number of scales and dimensionality did not make much of a difference in the final performance. For instance, 1:2:128 and 1:4:128 scales performed almost equally well, the former, however, having higher dimensions. Since, 1:4:128 performed best with ‘db5’ mother wavelet across the majority of configurations, it was chosen for the further experiments with R-KDA.

As discussed previously for R-KDA, we only experimented with three values of $\sigma = 10^5, 10^6, 10^7$ with a combination of three values of $\eta = 10^{-3}, 10^{-4}, 10^{-5}$, in total nine different parameter settings. The results suggested that $\eta = 10^{-4}$ performed better across the majority of configurations for CWT.

Table 5.1 presents results in EER for CWT+DLDA and CWT+R-KDA with both the classifiers for all configurations of BioSec.Lab PPG database. It can be observed that the best performance does not occur with the same feature reduction technique, classifier or parameters across different settings. The simple explanation is that training and testing data are not the same across all configurations. Therefore, optimal parameters for minimal training error and good generalization are also different for each configuration. In addition, and as discussed previously, feature reduction techniques and classifiers are also not fine tuned, therefore they are not guaranteed to produce consistent results across different configurations. Since our primary objective is only to evaluate capability of PPG as biometrics and not on finding out the best classifier or feature reduction technique, only the best results across all combinations of reduction techniques and classifier are further used for comparisons.

Fig. 5.1 shows comparison of best results for all configurations for both feature reduction techniques together with CWT without feature reduction. For the single session, performance as good as 3.57% EER was achieved. Since the single session has training and testing data from same session, it is not surprising. Within the short time lapse of at least 30 minutes, EER increased up to 29.19%. Doing exercise further

increased EER up to 40.69%. Having longer time differences of on average of 36 days, between two sessions had EER of 31.60%. By comparing long time-lapse with short time-lapse, we can say that the amount of time difference between training and testing does not degrade the performance of PPG biometrics by large amount. This observation is further verified by three sessions long time-lapse configurations, where variance in performance is small across different time-difference. Hence, it can be said that PPG has good permanence property but poor robustness when subjected to exercise.

It should be noted that, for three sessions long time-lapse configurations, time differences between the first and second session was on average 36 days while it was 128 days between the second and third session. Despite that, performance when training with the first session and testing with the second session is worse compared to training on the second and testing on the third. This is counter-intuitive, as generally with longer-time differences one should expect more degradation in performance. However, there are multiple factors which determine the performance rather than just time differences. For example, mental and physical states under which PPG signals were recorded, recording environment, quality of signal, and of course tuning of the system. Therefore, considering all those factors either controllable or uncontrollable, these results should only be regarded as *empirical results*.

Another observation that can be made from Table 5.1 is that Random Forest performed better compared to SVM across all configurations except for the short time-lapse. This can be due to over-fitting in SVM. Another reason could be of residual errors and other noises, which could deteriorate the performance of SVM while random forests having been trained using bootstrapping, randomness, and having ensemble of trees are more resistant to noise.

Fig. 5.1 also compares results with no feature reduction. For no-reduction also best results between SVM or Random Forest is taken for comparison. For all configurations, having feature reduction either DLDA or R-KDA always performed better than having no reduction. This asserts the belief that feature reduction in the case of high dimensions helps improve the performance. As discussed in Chapter 1, it should be noted that EER alone as a performance metric is not sufficient. The Receiver Operating Characteristics (ROC) need to accompany for deciding performance for application specific requirements. ROC curves for CWT results for all configurations are included in Appendix A.

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

1DMRLBP extracts micro textures surrounding each sample at multiple resolutions. Histogram operation afterward emphasizes the frequency of these patterns in a PPG segment. Feature vector generation through 1DMRLBP mainly depends on 5 parameters that are 1. Number of points to consider for each sample (p), 2. How far from the center sample, patterns start getting extracted (d) 3. Window length for the histogram, 4. Shift of the window (Δ) and 5. Leeway (ϵ). Details on how 1DMRLBP feature vector is generated can be found on Chapter 4.

To find out the best parameters of 1DMRLBP for the PPG biometrics, various experiments were con-

Table 5.2: Results with 1DMRLBP and DLDA. Columns represent set of parameters for 1DMRLBP, Rows are different database configurations for evaluation, and Cell values are performance in EER for each classifier. Best Results are highlighted in green color. Multiple Resolution are read respectively.

Number of Points (p)	(4,4)	(4,4)	(5,5)	(4,4)	(2,2)	(4,4,4,4)	(4,4,4,2,2)
Distance (d)	(1,10)	(10,20)	(10,20)	(10,20)	(10,20)	(1,10,10,20)	(1,10,10,20,10,20)
Window Length	(100,100)	(50,100)	(50,40)	(50,40)	(50,40)	(100,100,50,100)	(100,100,50,100,50,40)
Shift Δ	(0,0)	(10,0)	(10,20)	(10,20)	(10,20)	(0,0,10,0)	(0,0,10,0,10,20)
Feature Vector Dimensions	512	1792	10240	2560	160	2304	2464
Single Session	Random Forest	13.48%	9.07%	9.75%	8.01%	8.30%	7.35%
	RBF SVM	9.68%	5.95%	6.67%	6.33%	7.05%	6.50%
	Random Forest	36.05%	28.07%	29.38%	30.04%	31%	33.08%
Short Time Lapse		32.20%	29.09%	31.03%	30.90%	30.91%	31.00%
		Random Forest	37.90%	44%	42.25%	45.38%	41.15%
		RBF SVM	38.14%	47.50%	47.50%	45%	48.68%
Two Sessions Long Time Lapse		Random Forest	41.29%	37.76%	38.64%	39.52%	36.86%
		RBF SVM	42.10%	41.74%	43.24%	40.54%	41.19%
		Random Forest	46.06%	43.75%	42.61%	43.75%	47.95%
Three Sessions Long Time Lapse	Train on 1 st Test on 2 nd	RBF SVM	50%	42.03%	43.75%	47.92%	43.75%
		Random Forest	40.34%	41.34%	36.74%	40.04%	41.86%
	Train on 1 st Test on 3 rd	RBF SVM	40.83%	41.26%	43.01%	37.50%	37.50%
	Train on 2 nd Test on 3 rd	Random Forest	26.71%	30.14%	38.60%	40.68%	39.35%
		RBF SVM	27.50%	31.25%	33.34%	35.40%	31.66%
							31.25%

Table 5.3: Results 1DMRLBP with R-KDA, regularization was set to $\eta = 0.0001$, best results are highlighted in green.

	Points (p)	(4,4,4,4)		(4,4,4,2,2)	
	Distance (d)	(1,10,10,20)		(1,10,10,20,10,20)	
	Window Length	(100,100,50,100)		(100,100,50,100,50,40)	
	Shift Δ	(0,0,10,0)		(0,0,10,0,10,20)	
	Feature Vector Length	2304		2464	
	Kernel Width σ	1.00E+07	1.00E+06	1.00E+05	1.00E+07
					1.00E+06
					1.00E+05
Single Session	Random Forest	7.48%	8.40%	7.26%	7.19%
	RBF SVM	6.78%	7.14%	5.20%	6.84%
					7.14%
Short Time Lapse	Random Forest	31.73%	31.78%	34.92%	32.57%
	RBF SVM	30.90%	30.27%	29.09%	29.36%
					29.09%
Across Exercise	Random Forest	40.59%	42.09%	41.10%	42.33%
	RBF SVM	45.38%	45.45%	46.57%	47.50%
					49.17%
Two Sessions	Random Forest	40.25%	43.22%	37.76%	41.69%
	RBF SVM	40.54%	42.36%	43.24%	43.24%
					43.24%
Long Time Lapse	Random Forest	43.93%	45.46%	41.76%	41.15%
	RBF SVM	45%	43.75%	43.75%	45.83%
					43.75%
Three Sessions Long Time Lapse	Train on 1st, Test on 2nd	39.36%	39.50%	34.18%	39.39%
	Random Forest	43.75%	41.25%	37.50%	32.50%
	RBF SVM	29.33%	26.56%	28.50%	30.87%
	Train on 2nd, Test on 3rd	31.25%	31.25%	31.25%	31.25%
	Random Forest	43.75%	41.25%	37.50%	37.50%
	RBF SVM	31.25%	31.25%	31.25%	31.25%

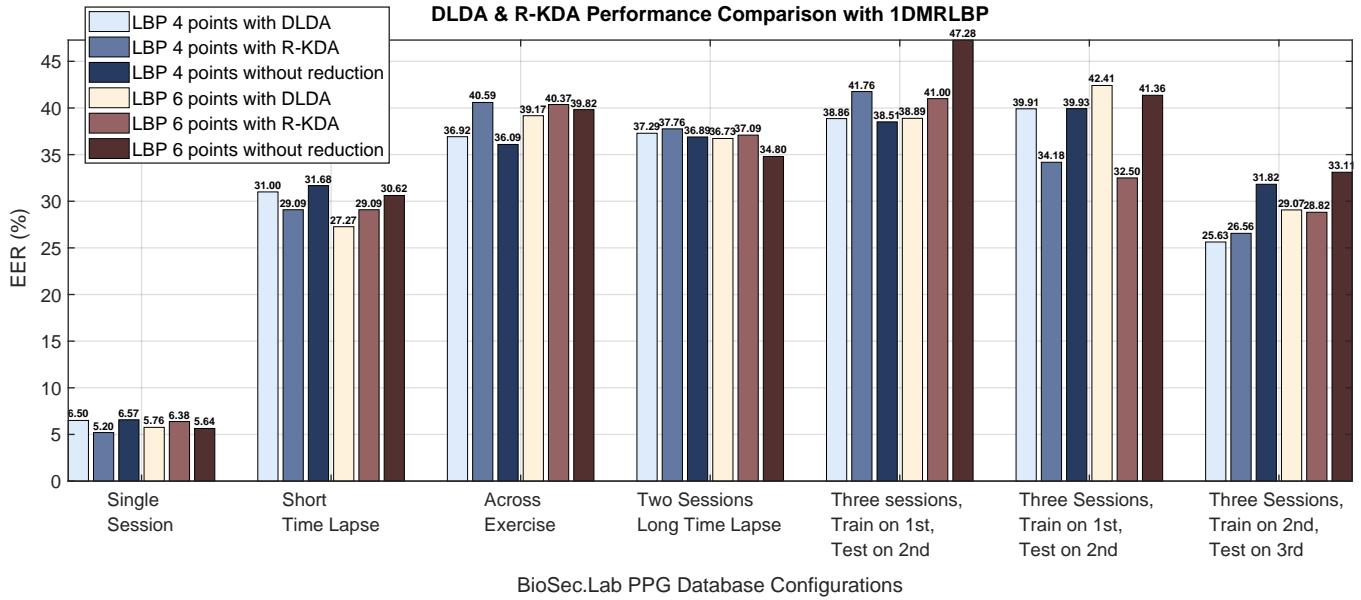


Figure 5.2: 1DMRLBP performance comparison by taking best result each feature reduction technique

ducted with DLDA as reduction technique. Table 5.2 presents results for these experiments. In Table 5.2, multiple resolutions are read respectively. For instance, when $p = (4, 4)$, $d = (10, 20)$ then, first LBP features are extracted with $p = 4, d = 10$, then with $p = 4, d = 20$. After that for the histogram operation, similarly the window and shift are applied respectively. An example of calculation of feature vector can be found in Chapter 4. During experiments, it was found that the majority voting performed better than constructing one feature vector from past n segments. Therefore, n was simply set to 1 in scheme mentioned in sec. 4.2.2.3.

During experiments, it was discovered that leeway parameter $\epsilon = 0.001$ performed better compared to having no leeway. This is also consistent with the findings of [62]. Hence all results in Table 5.2 are with $\epsilon = 0.001$. Same as CWT, here also best results do not occur with same parameters for each configuration. The same reasoning as CWT also applies here. It can further be observed that having multiple resolutions that is considering more points at different parameters improve results compared to a lesser number of points. For example $p = (4, 4, 4, 4)$ performed better than just $p = (4, 4)$. It was also found that, considering only 4 or 6 resolutions were sufficient for good performance. Taking more resolutions beyond that did not improve performance.

Since, the last two settings in Table 5.2 that are with $p = (4, 4, 4, 4)$ and $p = (4, 4, 4, 4, 2, 2)$ performed best across the majority of configurations, they were chosen for the further experiments with R-KDA. Same as CWT, here also R-KDA experiments were conducted with only the combination of $\eta = 10^{-3}, 10^{-4}, 10^{-5}$ and $\sigma = 10^5, 10^6, 10^7$. $\eta = 10^{-4}$ was found to perform better across a majority of configurations. Table 5.3 presents results of R-KDA with $\eta = 0.0001$ for all three values of σ , for both 4 points 1DMRLBP and 6 points 1DMRLBP settings, for all configurations and both classifiers. Here also, the best results do not occur with same parameters for all configurations. In addition, it can be observed that for both DLDA and

R-KDA, Random Forest performs better than SVM for the majority of the configurations. Same reasoning of over-fitting and noise as mentioned in CWT results also apply here.

Fig. 5.2 compares the best results found for DLDA, R-KDA, and no-reduction for all configurations, where LBP 4 stands for $p = (4, 4, 4, 4)$ and LBP 6 is $p = (4, 4, 4, 4, 2, 2)$ 1DMRLBP setting. For no-reduction results the best result was taken between Random Forest and SVM for comparison. However, with no-reduction the best results always occurred with Random Forest while RBF SVM in high dimensions struggled to give a good classification. The important thing that can be taken from Fig. 5.2 is that, no-reduction performed as good as DLDA or R-KDA reduction. In a few cases, it performed even better than DLDA or R-KDA, for example with the two session long time-lapse. It can be attributed to the fact that, 1DMRLBP feature vector despite being high dimensional, it is very sparse by nature and Random Forest training is not affected in that case while DLDA or R-KDA requires sufficient number of training samples and tuning. For single session, best EER of 5.20% was achieved while 27.27%, 36.09%, 34.80% for short time-lapse, across exercise and two sessions long time-lapse configurations respectively. Analogous to CWT results, here also it can be seen that for three sessions, training on the first session, testing on the second session has higher EER compared to training on the second, testing on the third, supporting the belief that there are several uncontrollable factors that also affects that performance than just time-difference. ROC curves for each configuration are included in Appendix A.

5.2.3 Results with Auto-Correlation (AC)

Auto-correlation measures the similarity of a signal with its own shifted version. For the auto-correlation features, pre-processing was done differently as it did not require segments to be synchronized with each other. First, PPG signal was blindly segmented into overlapping windows. Following this, the AC was taken for each signal window. Outlier AC windows were discarded using the same method described in Chapter 4. AC windows were then projected onto subspace using DLDA or R-KDA such that classes were well separated with each other. At the end, projected segments were classified using SVM or Random Forest.

AC based features depend on mainly 3 parameters. 1. Window length used to segment signal. 2. Overlap between each window 3. Number of AC lags. For each of these parameter, exhaustive search was carried out to find best combination. Every possible combination of following values was used in exhaustive search with DLDA and both classifiers for every configuration of BioSec.Lab PPG database,

Window Length (in seconds)	:	4, 5, 6, 7, 8
Overlap (in seconds)	:	1, 2, 3, 4
Number of AC lags (in samples)	:	25, 30, 40, 50, 60, 70

Through the search, it was found that none of the combinations of parameters performed as good as 1DMRLBP or CWT. Further, there was no single set of parameter values that performed well across all configurations. For instance, the first 70 lags performed better than other values for the single session, while the first 40 lags performed better for the across exercise. It is not surprising as for relaxed state data, PPG

Table 5.4: Results with AC as features. Columns represent feature reduction technique with classifier. Rows are different configurations from database. Cell values are the performance in EER. For, R-KDA, results are from three different values of parameter σ . The best Results are highlighted in green.

		DLDA			R-KDA with $\eta = 1e - 5$		
		Random Forest	RBF SVM	Random Forest	1.00E+07	1.00E+06	1.00E+05
Single Session	23.43%	24.51%	29.26%	25.60%	24.04%	29.65%	27.32%
Short Time Lapse	46.81%	43.64%	45.83%	47.58%	46.56%	45.45%	45.45%
Across Exercise	43.61%	45.50%	47.91%	53.14%	47.32%	50.00%	47.50%
Two Sessions Long Time Lapse	40.13%	45.95%	46.85%	48.15%	43.33%	45.95%	47.50%
Three Sessions Long Time Lapse	Train on 1st, Test on 2nd	44.41%	46.67%	50.34%	45.63%	50.93%	51.60%
	Train on 1st, Test on 3rd	43.41%	50.00%	40.54%	43.80%	42.25%	47.19%
	Train on 2nd, Test on 3rd	43.01%	38.24%	43.75%	42.86%	40.54%	41.80%

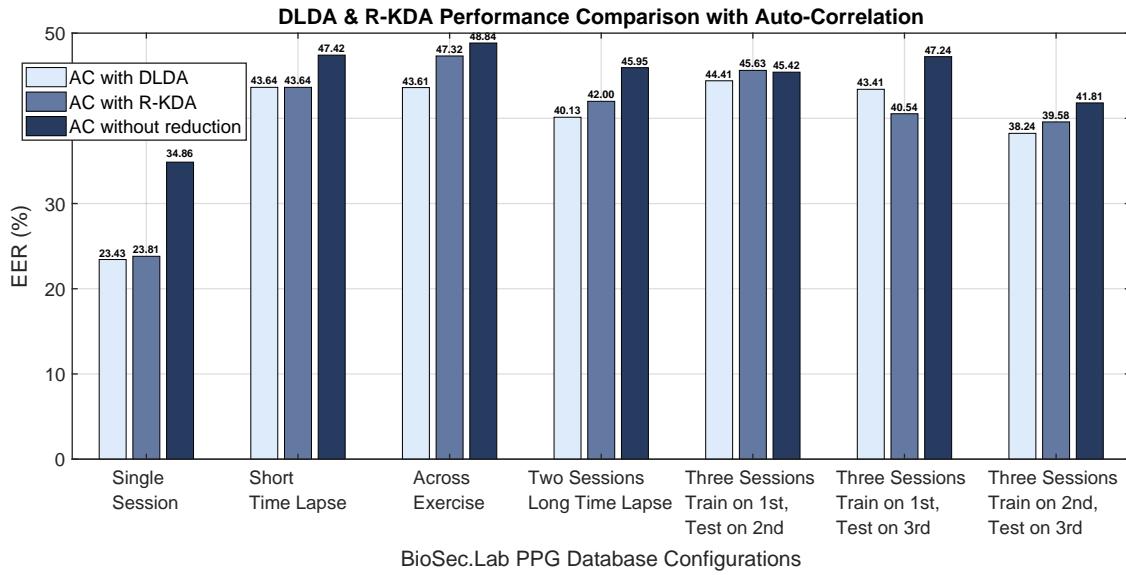


Figure 5.3: AC performance comparison by taking best result each feature reduction technique

beat is longer while for exercise it shrunk in length. Hence, for the across exercise selecting only the first few lags generated better similarity measure compared to the relax data. For ECG, it was observed that selecting only the first few lags equal to the length of QRS complex produced good results as QRS complex was found to be robust to changes [92]. Unfortunately, that is not the case with PPG. Hence, no single value of the ‘number of lags’ provided consistent results. Additionally, the problem was further aggravated when window length and overlap factors were also counted. Out of all the combinations of the three parameters described, window length of six seconds with an overlap of three seconds and the first 60 lags were finally chosen for further experiments with R-KDA. This set of values was found to be performing well compared to other values in general across all configurations. For brevity, only results for these selected parameters are included in this thesis.

Table 5.4 presents results of AC features with DLDA and R-KDA. It can be seen that, even for the single session, the best EER is 23.42%. Analogous to CWT and 1DMRLBP, the best results do not always occur with same set of parameters or classifier. Random forest performed better on majority of the configurations. For R-KDA, regularization parameter $\eta = 10^{-5}$ was selected as it performed better than other values. Smaller value of η compared to 1DMRLBP and CWT can be associated with the fact that, AC features are smaller in dimensionality and requires less regularization. In addition, DLDA performed better than R-KDA in the majority of the configurations, which again can be due to smaller dimensions and improper tuning of R-KDA. Fig. 5.3 compares the best results taken for each reduction technique together with no reduction for all configurations. It can be noticed that having feature reduction helped improve performance compared to no-reduction. The best EER of 43.64%, 43.16%, 40.13% was achieved for the short time-lapse, across exercise and two sessions long time-lapse configurations respectively. In addition, for three sessions configurations, the best EER for training on the first and testing on the second is higher compared to training on the second and testing on the third, which is again consistent with results of CWT and 1DMRLBP. ROC curves for AC

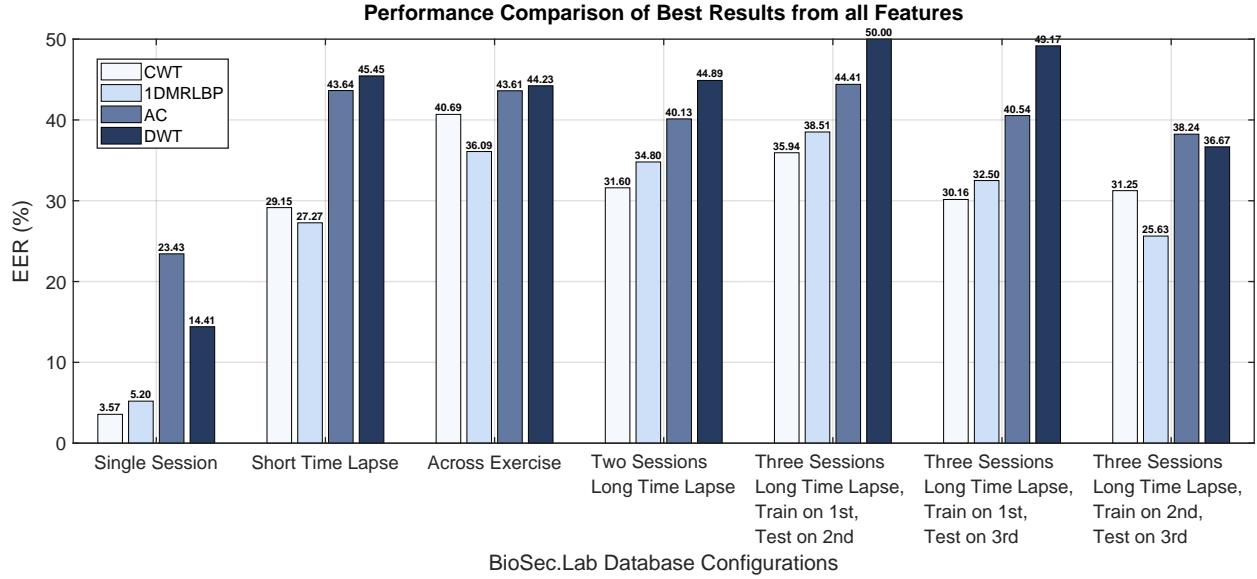


Figure 5.4: Best Performance Comparison for all Features

results are included in Appendix A.

5.2.4 Comparison of results from all features

After finding out the best performing combination of feature reduction, classifier, and various parameters for each feature extraction techniques, it is desired to compare results from each of these features for all configurations. Comparing this way would give us an idea of the best achievable performance for PPG biometrics by any system under different circumstances. Fig. 5.4 presents this comparison where results from Discrete Wavelet Transform (DWT) based system are also included. The DWT system was adapted from paper [46]. Authors of [46] showed promising results with DWT based system for Capnobase database. In particular, system of [46] was chosen for comparison of our results because presented results in the paper with DWT were state of the art, in the sense that sample size used for results was the largest in literature and presented results were also best amongst all papers in literature.

In the DWT system, features were processed in two steps. First the Kolmogorov Smirnov test (KS-test) was used to remove redundant features. Following this, remaining features were reduced to dimensionality of 10 using Kernel PCA. Classification was carried out using the Nearest Neighbor Data Descriptor (NNDD). DWT features were extracted using ‘db4’ mother wavelet using four levels of decomposition. For fair comparison, same pre-processing and evaluation framework as CWT, 1DMRLBP, and AC was used to generate results for DWT system.

From Fig. 5.4, it can be observed that CWT performed best four times while three times for 1DMRLBP out of all seven configurations. AC and DWT based feature performed worse for all configurations relatively. DWT based system performed worst among all in general which can be attributed to three factors. First,

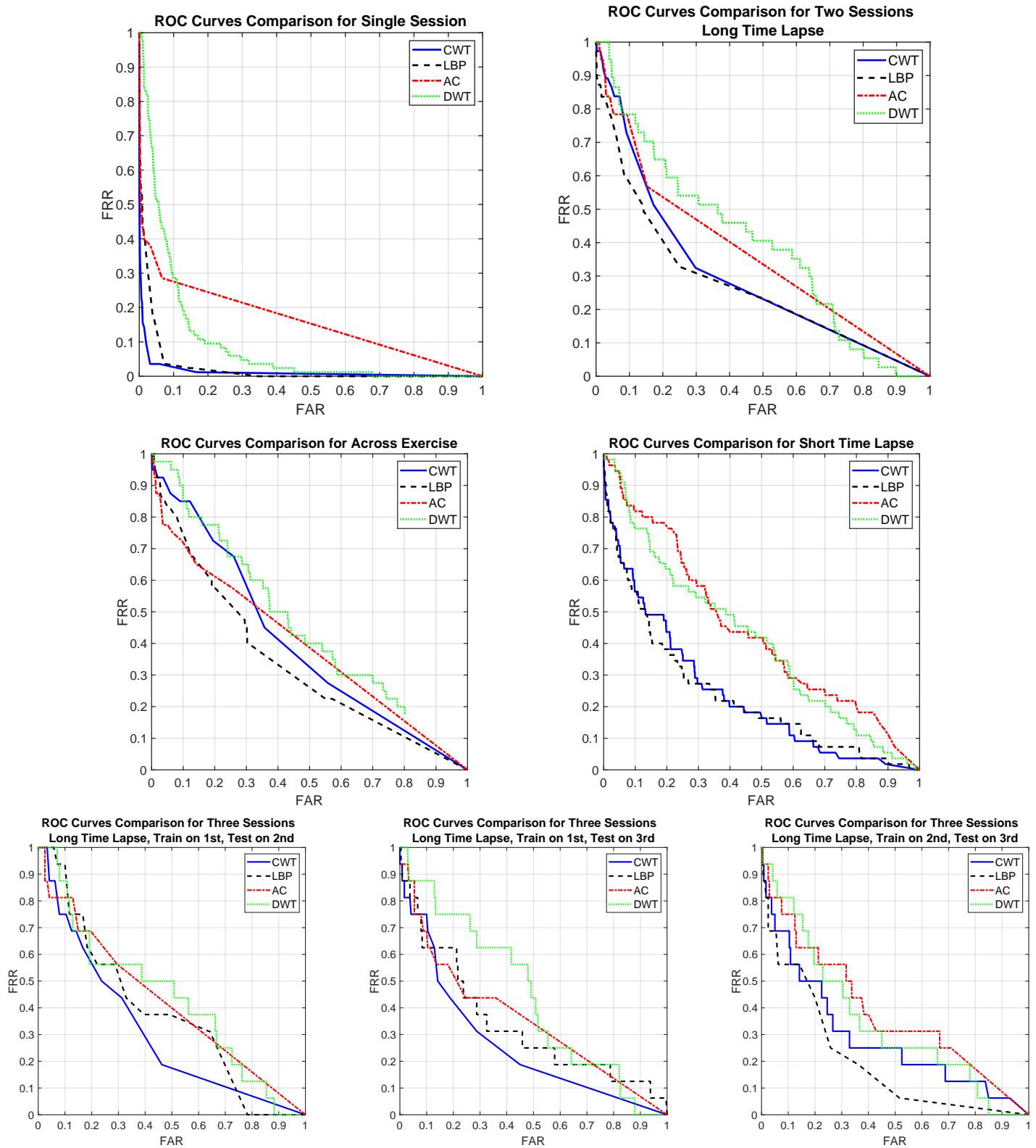


Figure 5.5: Receiver Operating Curves comparisons for best results from all features

KS-test based redundant features removal can be regarded as a filter approach of feature selection which does not optimize for better classification accuracy while selecting features. Moreover, in multiple session scenario, it fails to account for possible changes in data during training while selecting features. Therefore, despite producing satisfactory results with the single session, DWT system performed the worst for rest of the configurations. Secondly, kernel PCA is an unsupervised reduction technique which does not emphasize class separation. Lastly, unlike Random Forest and SVM, NNDD is more sensitive to noise and outliers. The DWT based system in three sessions long time-lapse, training on the first and testing on the second has a higher EER compared to training on the second and testing on the third, which is again consistent with results from the rest of the features.

Bad performance of AC based features compared to CWT and 1DMRLBP can be due to the fact that auto correlation features are less descriptive in the sense that it is a similarity measure with signal itself. Contrary to that in fact along with AC, a strong cross-similarity is also needed for better performance as PPG signal changes with time, and one's physical and mental state. Hence, it struggles to serve as good features as it does not have strong pattern extraction or detailed representation like 1DMRLBP or CWT.

It should be recalled that one limitation of direct comparison among results from different configurations is that the number of subjects for each configuration are not equal. Therefore, it is not completely fair comparison. Currently, there is no study on the effect of sample size on performance of PPG biometrics. Hence, comments on how much fair these comparisons are, can not be made. However, EER which depends on FAR and FRR are in fact normalized performance metrics. Hence, drawing conclusions from these comparisons are not completely invalid.

From Fig. 5.4, it can be noted that EER of 3.57% was achieved for the single session configuration which included 84 subjects. However, within the short-time lapse of 30 minutes, it increased up to 27.27%. On the other hand, in the single session configuration despite training and testing data coming from the same session or same continuous recording, the good performance with a sample size of 84 subjects shows that PPG does in fact carry distinct information for each individual within a short time-frame. If PPG signal was completely random or indistinct then this error would have been much higher particularly when having sample size as large as 84 subjects. Hence, distinctiveness property of PPG cannot be denied. However, because of various factors such as time lapse and one's physical and mental state, PPG shows greater variability causing degradation in performance.

When considering only the best time lapse results from Fig. 5.4, we can see that EER varies between 27.27% to 35.94%. Further, If results for the train on the first and test on the second configuration is discounted, as data for that seems anomalous, then in fact EER only varies between 27.27% to 31.60% for different amount of time lapse from minimum of 30 minutes for short time-lapse to on average 164 days of time lapse between the first and third session of the three sessions long time-lapse. In practice, this variation in performance can further be reduced by fine tuning the system. Hence, we can conclude that if a system is properly trained and tuned, then PPG biometrics has good permanence property. However, with exercise EER rises up to 36.09%, and is therefore less robust towards physical stress.

Another thing to consider is that, EER does not provide insights for application specific performance.

For instance, for some application one could only afford to have False Acceptance Rate (FAR) of a maximum 10%. In such cases, performance can better be estimated from Receiver Operating Characteristics (ROC). Fig. 5.5 shows ROC curves for best results from all features for each configuration. It can be seen that except for the single session, False Rejection Rate (FRR) varies approximately between 50% to as high as 70% at 10% FAR. The significance of 50% FRR is that every other trial would be rejected from a genuine user. Based on the desired application, this can be acceptable or unacceptable. For example, considering 20 PPG segments are used for the verification each time, the same as in this thesis then two genuine trials in case of 50% FRR would approximately take only 40 seconds assuming noise-free signal acquisition. This can be even faster than typing lengthy complicated passwords for a few people. Hence, based on the required verification speed, accuracy, and other operational factors, PPG can be useful for some applications. FRR increases further at 1% FAR varying approximately between 78 – 100% for all configurations, except for the single session. Therefore, PPG biometrics system as described in this thesis may not be suitable for high security applications with quick verification demands.

In addition, when looking at single session ROC, it can be seen that even at 1% FAR, FRR is approximately between 10 – 15%, which can be suitable for high security applications. Hence, to make use of PPG as biometrics for high security applications, one can also come up with single session like evaluation setting to have good performance, which in fact is the scheme of so called ‘continuous authentication’. In continuous authentication scheme, instead of focusing much on variations in PPG, system can rather capitalize on better performance of single session like results. More discussions on this are presented in Chapter 6 of the conclusion and future directions.

5.3 Results with Capnibase Database

Capnibase database has been used by a few recent papers on PPG biometrics [46], [47], [49]. Database contains PPG signals recorded at 300 Hz from 42 subjects in a controlled environment. Since the database was recorded in a controlled environment with the purpose of developing respiratory rate estimation algorithm, signals from this database show greater respiratory induced variations. Hence, it is also desired to assess the performance of PPG biometrics under such circumstances. It should be noted that the database only contains signals recorded in single session. All signals were resampled at 100 Hz to match with BioSec.Lab PPG database.

For simplicity’s sake, experiments with only DLDA have been performed on this database. Further, parameters for all the feature extraction techniques were kept same as BioSec.Lab PPG database experiments. That is for CWT features, ‘db5’ mother wavelet with 1:4:128 scales was used. Similarly, for 1DMRLBP 4 points parameter setting with $p = (4, 4, 4, 4)$ and 6 points parameter setting with $p = (4, 4, 4, 4, 2, 2)$ were used. For Auto-correlation window size of six seconds with three seconds overlap and the first 60 lags were used for extracting features. Training, testing, and pre-processing were also kept same as BioSec.Lab PPG database.

Fig. 5.6 presents results from CWT, 1DMRLBP, and AC on Capnibase database with DLDA as feature

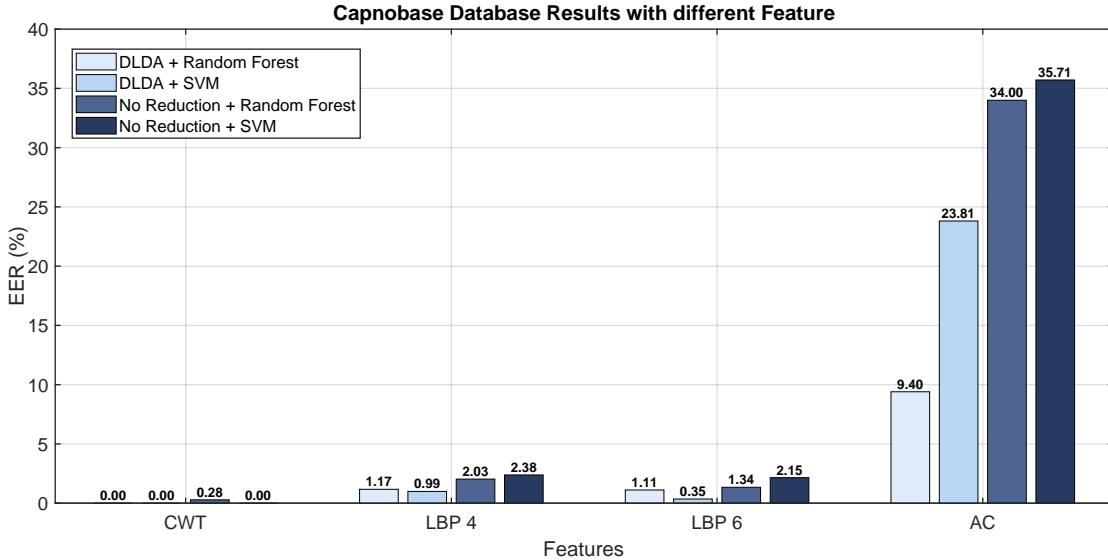


Figure 5.6: Comparison of results with Capnibase Database for different features

reduction and Random Forest and SVM as classifiers. Best EER of 0.00% and 0.35% was achieved with CWT and 1DMRLBP respectively. It can be noticed that auto-correlation performed severely worse for capnibase. Reason for that is signals from capnibase database are affected by respiration induced noise. Secondly, auto-correlation parameters were kept the same as the BioSec.Lab PPG database and were not tuned for Capnibase database. In addition, by comparing AC results of fig. 5.6 with the single session results from Fig. 5.3, it can be noticed that results from both databases for auto-correlation are in fact in a similar range. Fig. 5.6 also contains results without any reduction for comparison and for which EER is higher compared to DLDA, affirming that ‘LDA’ based feature reduction helps improving classification.

Fig. 5.7 compares the Receiver Operating Characteristics (ROC) for the best results from each feature, where results from DWT based system described earlier is also included for comparison. DWT system was implemented from [46] which used the same Capnibase database. Our system performed better at 0.00% EER compared to 4.30% with DWT based system on the same database. From Fig. 5.7 it can be observed that even at 1% FAR, FRR is much lower. From this observation, it can be concluded that PPG is feasible for high security biometrics applications under controlled environment and if a single session like evaluation scheme is used.

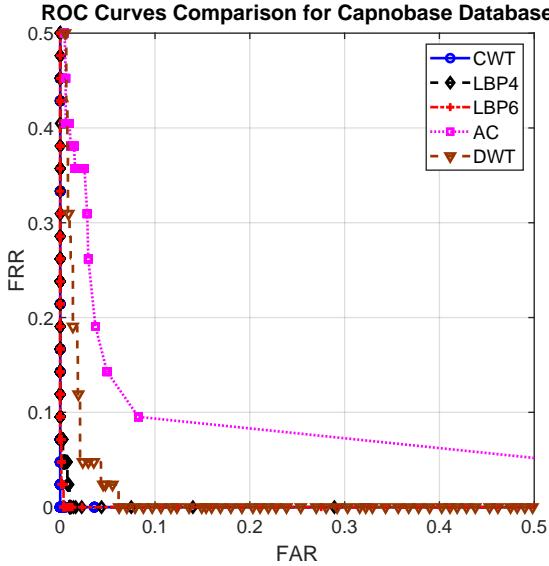


Figure 5.7: Capnibase ROC Comparisons

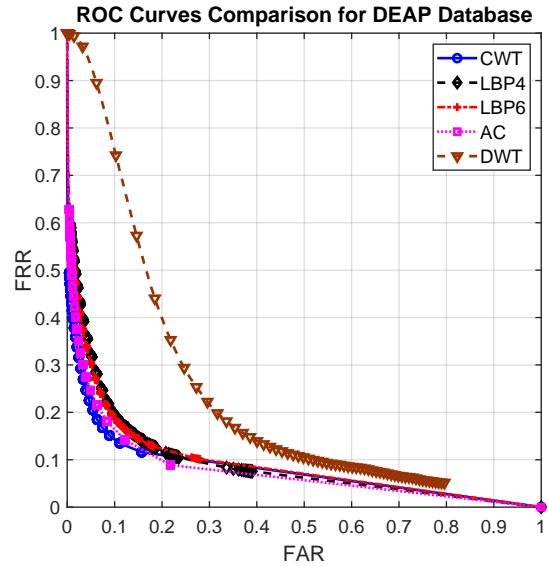


Figure 5.8: DEAP ROC Comparisons

5.4 Results with DEAP database

DEAP database contains one minute long PPG recordings at 128 Hz from 32 subjects under 40 different emotional stimuli. It is one of the benchmark databases widely used in affective computing. Details on how signals were collected along with the mechanism of emotions stimulation can be found in Chapter 3. It should be noted that, each one minute recording took place with five seconds of baseline recording and two seconds of additional time difference between them. Further, participants were given a short time break after 20 recordings. Therefore, from a biometrics perspective along with emotional stress, a PPG recording would also exhibit short time-lapse effect.

Main objective behind experiments with DEAP database was of assessing the effect of emotional stress on performance of PPG biometrics. Therefore, during experiment, the training set consisted of 80% of the signal from one of the forty emotions and test set included rest of the 20% signal from the same emotion and other 39 emotions for each subject. To elaborate more, for each subject 40 genuine trials and $31 \times 40 = 1240$ imposter trials were carried out to calculate FAR and FRR. This was repeated 40 times, each time training with different emotion and the average results are reported here. Prior to pre-processing, signals were resampled at 100 Hz to match with BioSec.Lab PPG database settings. Similar to other two databases, majority voting from first 20 PPG segments from each signal was used for decision making during testing.

Since this database requires to deal with in total 40 minute long PPG signal each time for each subject for 40 iterations, it is computationally expensive. Therefore, only experiments with DLDA were carried out. Fig. 5.9 compares results of DLDA and no-reduction with both Random Forest and SVM classifier for each feature extraction technique. Parameters for each feature extraction technique were kept exactly same as BioSec.Lab PPG and Capnibase database. From Fig. 5.8, it can be seen that EER as low as 12.79% was

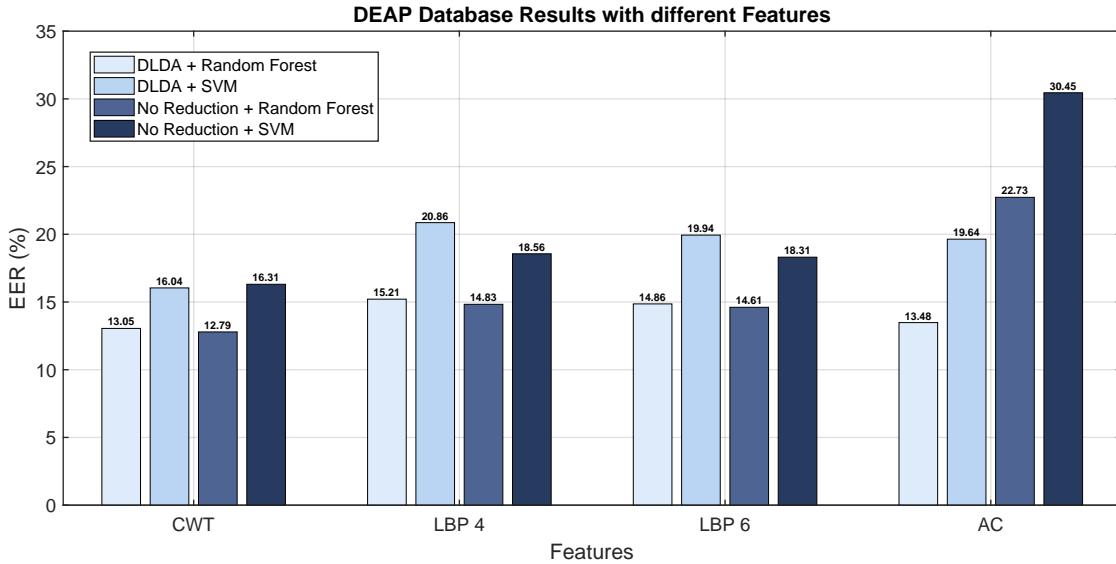


Figure 5.9: Comparison of results with DEAP Database for different features

achieved under emotional stress which is comparable to accuracy of 90.53% on the cross-session analysis reported by [37] on the same database with 20 segments. However, it should be noted that [37] only used data from 23 subjects out of 32 and the test set only consisted of 39 emotions from each person while in our case it consisted of data from all 40 emotions for each person.

From Fig. 5.9, it can be seen that having no-reduction performed slightly better than DLDA for CWT and 1DMRLBP features. It can be due to less number of training samples and high variability across 40 sessions because of which DLDA possibly struggled to find subspace that included invariant features across all sessions and provided better class separation. Nevertheless, the performance difference with DLDA is very small compared to no-reduction, which can be improved by fine tuning classifiers, using regularized DLDA (R-LDA) or by using R-KDA. Also, not surprisingly performance of 4 points LBP and 6 point LBP features are also almost same. The reason is that results are averaged over 40 iterations and variations in performance are settled. Another observation can be made is that Random Forest always performed better than SVM. That is also not surprising because across as many as 40 sessions, signals could change a lot and random forest is better suited for accounting variations. Auto correlation features also performed well but only with DLDA.

Fig. 5.8 compares ROC of best results from all features. Here, also DWT based system performed worse because of KS-test, Kernel PCA, and NNDD as explained earlier. It can be seen that at 10% FAR, FRR is low around 14%. In addition, even at 1% FAR, FRR is only around 35% meaning two out of three genuine trials would be accepted even at 1% FAR, which is remarkable. This suggests that the performance of PPG biometric system is less affected by emotional stress and the short duration of time-lapse.

Chapter 6

Conclusion and Future Work

6.1 Summary and Conclusion

This thesis started with primary objective of assessing capability of PPG as biometric modality. The idea was motivated by the concept of medical biometrics and its operational benefits over traditional biometrics traits. To establish any trait as biometric identity, one has to evaluate its distinctiveness, permanence, and robustness to changes. PPG being a physiological signal, it is important as it changes continuously due to various factors. Therefore, a comprehensive study conceiving all these factors was essential before making a conclusion about utility of PPG as biometrics. Despite that, during the literature survey, it was found that most of the studies in the past only focused on single session evaluation where training and testing data came from the same continuous recording. Also, results from each of these studies pointed towards different conclusions on the utility of PPG as biometrics causing further dilemma. In addition, the largest sample size for any previous study was limited to only 44 subjects. Despite the idea of using PPG as biometrics being floating around for the last 15 years, only around 20 papers were published in the past. One of the major hurdle was a lack of biometrics centric PPG database. Hence, to fulfill our goal of establishing PPG biometrics, we collected the largest PPG biometrics database named ‘BioSec.Lab PPG database’ comprising PPG signals recorded in four different configurations i.e. single session, short time-lapse, across exercise, and long time-lapse from as many as 84 subjects. Having different configurations assisted in evaluating the distinctiveness, permanence and robustness properties of PPG biometrics. In addition to BioSec.Lab PPG database, two other databases namely Capnibase and DEAP were also used to investigate the effect of respiration and emotional stress respectively.

Inspired by their success in ECG biometrics, experiments were performed with CWT, 1DMRLBP, and AC as features. DLDA and R-KDA were added as feature processors to improve class separability and reduce feature dimensions. Also, detailed care was taken during the classifier stage to further account for class-imbalances during training. To achieve the best performance with the designed system, effects of various parameters and hyper parameters were studied and tuned to get best possible results.

Best EER of 3.57% and 0.00% was achieved for single session configuration of BioSec.Lab and Capnobase database which affirmed the fact that PPG carried distinct information for each individual. During the permanence study, EER ranged between 27.27% to 35.94%, and was achieved for time difference of a minimum of 30 minutes, to as large as 164 days on average. An increase in EER compared to the single session was attributed to physical and mental changes along with other factors such as the recording environment, noise, system errors, and so on. Physical stress of exercise and mental stress of different emotions produced the best EER of 36.09% and 12.79% respectively.

ROC inspection revealed that despite a seemingly high EER across multiple sessions, PPG still could be useful for small scale low security applications. For instance, at 10% FAR, FRR ranged between 50 – 70% and was around 14% for multiple session configurations of BioSec.Lab PPG database and DEAP database respectively. Based on application requirements, this performance can be acceptable.

After analyzing all results and different factors, we can summarize the conclusions as follows:

- Distinctiveness: Having achieved low error results on large sample size, the distinctiveness of PPG cannot be denied. However, we feel that there is still need for intensive study to further reinforce this proposition.
- Permanence: Low variance in performance across time lapse of a minimum of 30 minutes to as large as 164 days on average proves that PPG biometrics has good permanence property.
- Robustness: Because of the extreme heart rate change, PPG biometrics is a less robust under intensive physical stress (exercise). However, experiments with mental stress (emotions) showed less degradation in performance which can be due to a lesser amount of heart rate variability. Hence, we can conclude that the robustness of PPG is dependent on the amount of heart rate variability.
- The application of PPG biometric system would largely depend on two factors which include: (1) Required authentication speed, and (2) Accepted amount of FAR. Lower FAR would lead to longer authentication time and vice-versa. From the results, we can conclude that PPG biometrics system as described in this thesis, can be useful for small scale low security application.

6.2 Future Work

6.2.1 Improvements

During this work, a large amount of time was spent finding out the best features extraction technique and related parameters. It was found that because of various reasons, some features and parameters worked better for some configurations. However, there was no single set of parameters that worked consistently good across all configurations. In the future, to address this problem and to have only one feature set that works well across all configurations, we can form a large feature vector by concatenating all best performing features from all configurations. Thereafter, to reduce dimensions, one can employ ‘Feature Selection’ methods to smartly select features that work well across all configurations.

Other idea is to completely avoid trouble of finding the best features and its parameters. This can be done using neural networks with some constraints. In the future, we plan to implement state of the art neural network based models developed for face recognition and adapt it for PPG biometrics. For example, ‘FaceNet’ model which along with feature extraction also performs feature selection intrinsically to provide maximum class separability.

Further improvement in performance can also be made by doing ‘continuous authentication’. As presented in this thesis, single session performance is much better compared to other configurations. In ‘continuous authentication’ scheme, a model is updated by authenticating the user after every predefined time interval. This way, the troubles of variations in signals can be avoided to some extent. In the future, we plan to develop such an algorithm with streaming PPG signals from wearable devices.

6.2.2 Other Directions

- To improve performance further for high security applications, multi-modal system together with ECG can also be developed. ECG and PPG both being cardiac signals are synchronized and provide all advantages of the medical biometrics.
- Remote-PPG has been used several times as anti-spoofing measures in face recognition literature. Recently, the idea of remote-PPG alone as biometrics is also surfacing. Hence, it would also be interesting to consider a multi-modal system of face recognition and remote-PPG.
- PPG also has been used as an anti-spoofing measure in finger print recognition. Given the flexibility in recording site and the small form factor of sensors, the idea of multi-modal system with finger print can also be investigated.
- Use of medical signals is also widespread for secure key generation in Body-Area-Sensor-Networks (BASN). Given the variability in PPG signals, it can be successfully used for the same.

Appendix A

ROC curves for BioSec.Lab Database

As discussed in Section 1.4, EER alone is not sufficient for performance insights and comparisons. ROC need to supplement it. Here, ROC curves for the best results in all three features and all configurations of BioSec.Lab database are included. An analysis of ROC can be carried out similar to the discussion laid out in Section 5.2.4.

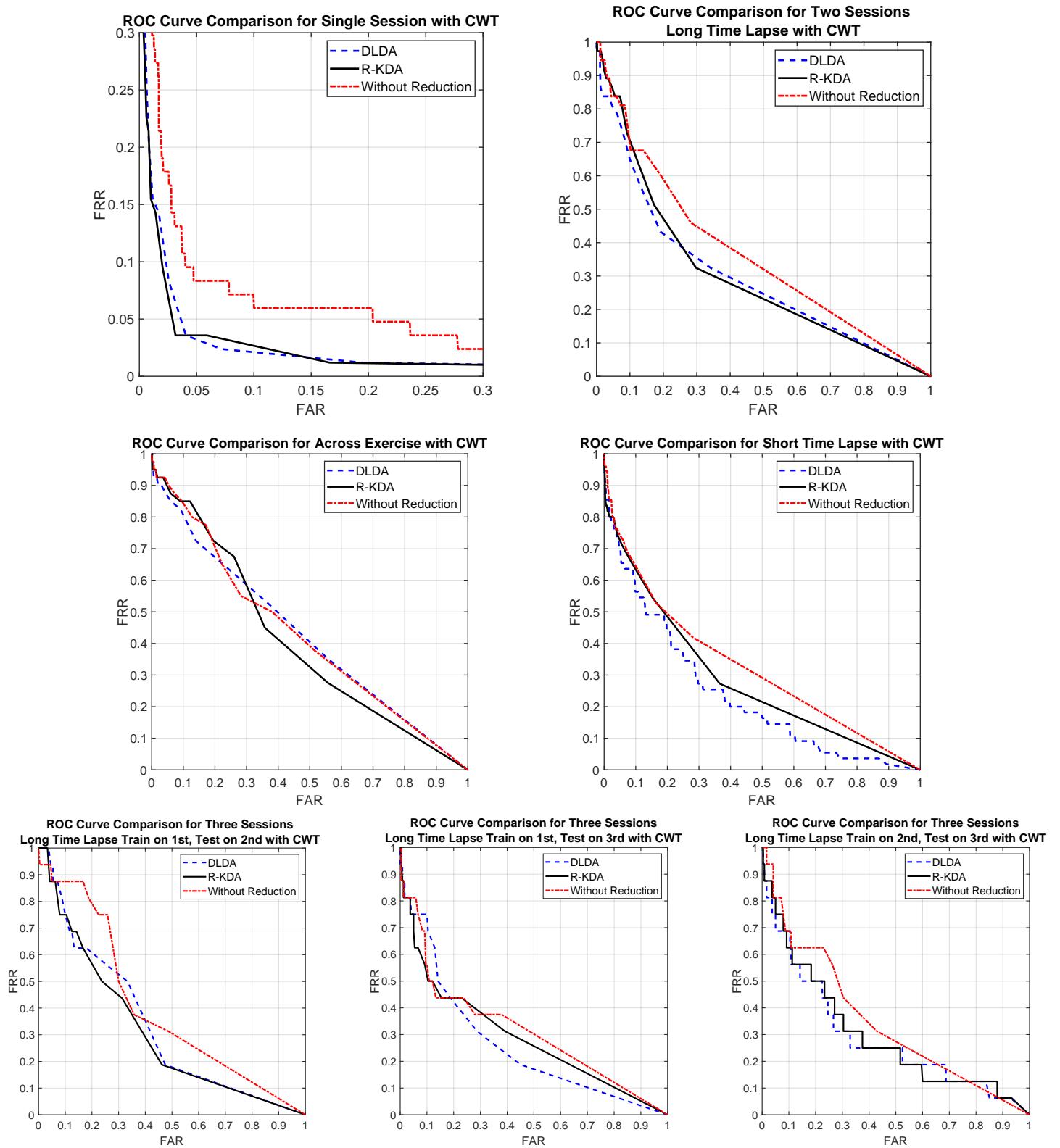


Figure A.1: Receiver Operating Curves for CWT

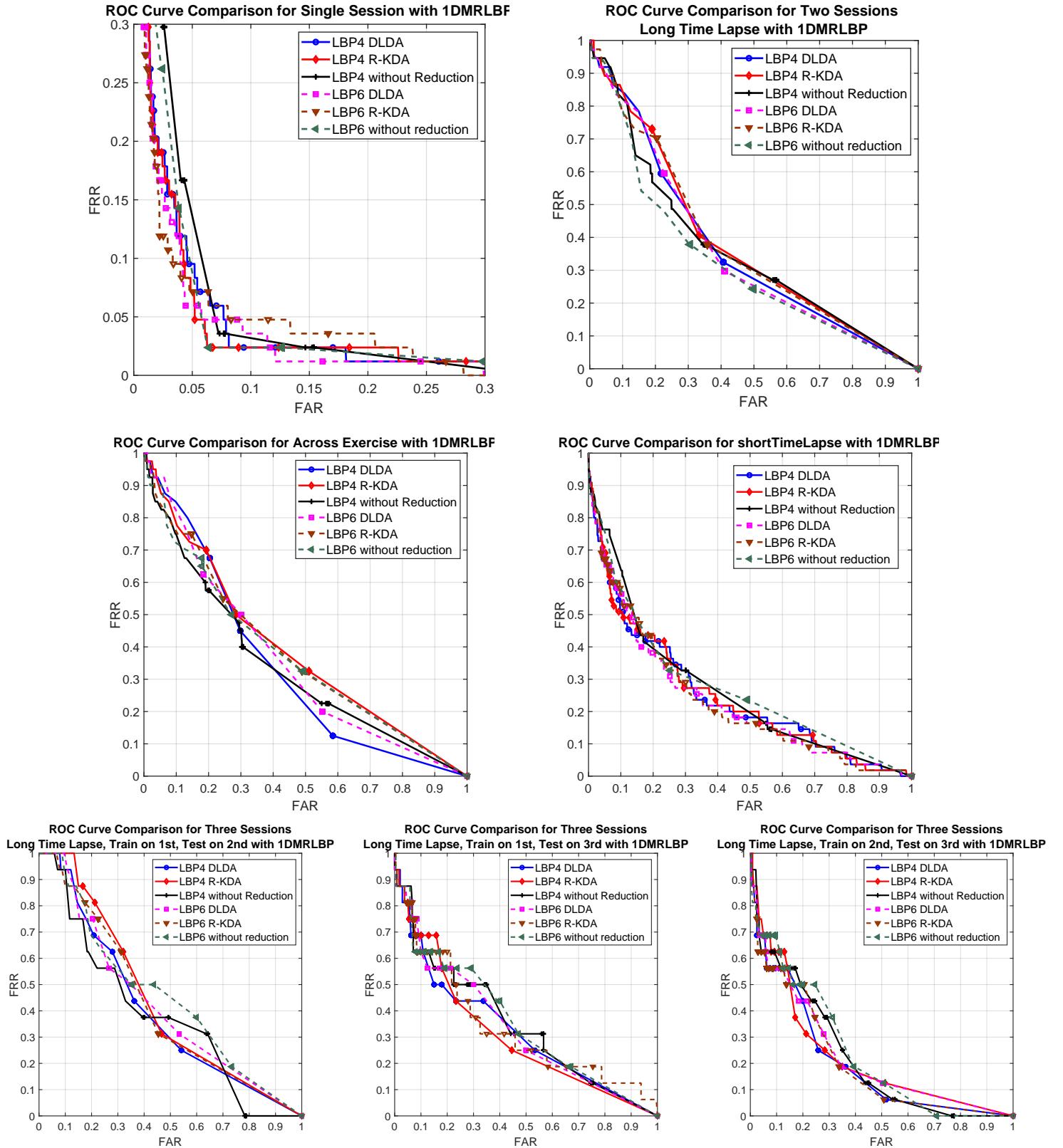


Figure A.2: Receiver Operating Curves for LBP

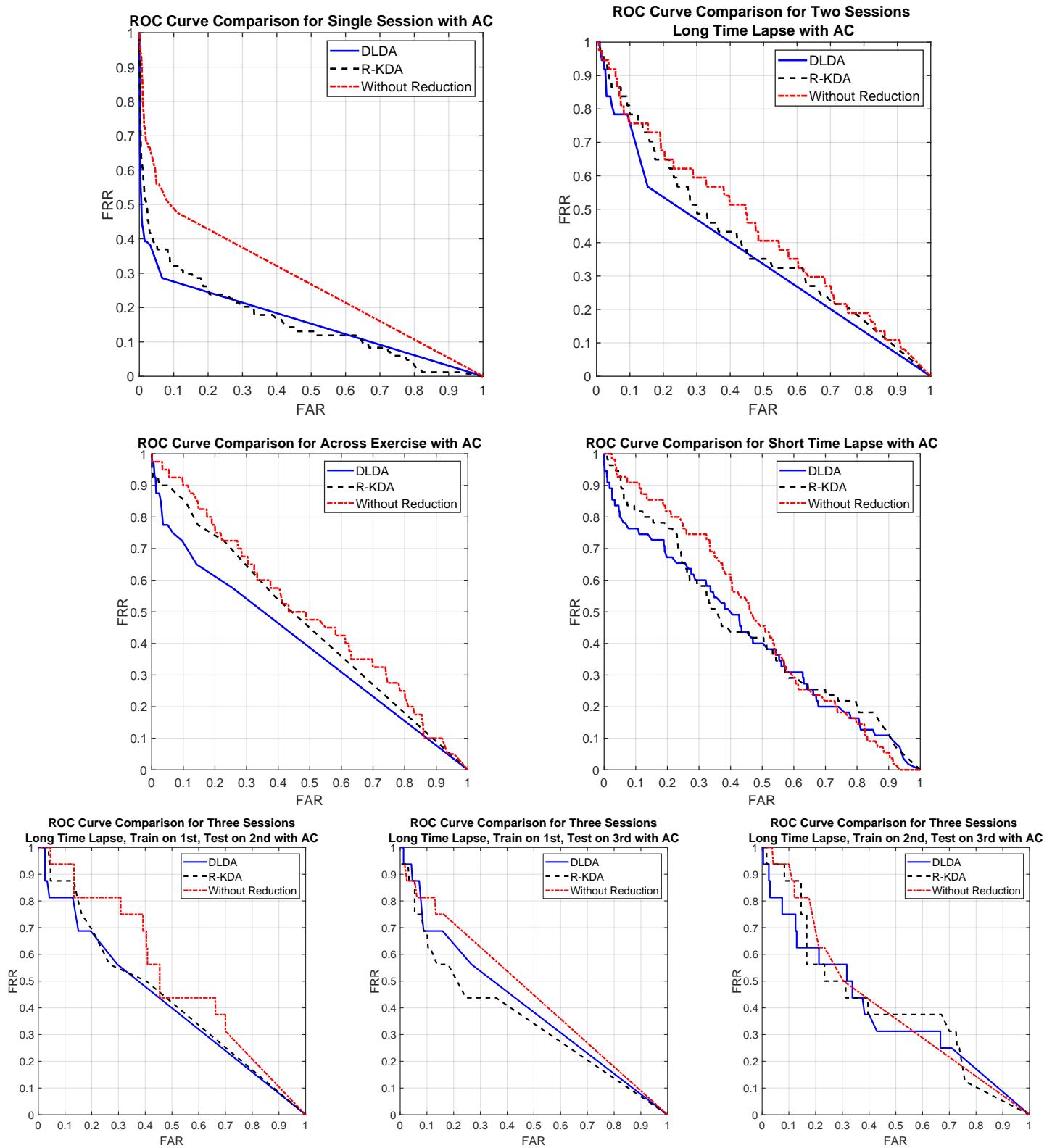


Figure A.3: Receiver Operating Curves for AC

Bibliography

- [1] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication", *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2021–2040, 2003, ISSN: 0018-9219. DOI: 10.1109/JPROC.2003.819611.
- [2] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities", *Pattern Recognition Letters*, vol. 79, pp. 80 –105, 2016, ISSN: 0167-8655. DOI: <https://doi.org/10.1016/j.patrec.2015.12.013>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865515004365>.
- [3] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition", *IEEE Access*, vol. 2, pp. 1530–1552, 2014, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2014.2381273.
- [4] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors—past and present", *Electronics*, vol. 3, no. 2, pp. 282–302, 2014, ISSN: 2079-9292. DOI: 10.3390/electronics3020282. [Online]. Available: <http://www.mdpi.com/2079-9292/3/2/282>.
- [5] E. M. Nowara, A. Sabharwal, and A. Veeraraghavan, "PPGSecure: Biometric presentation attack detection using photoplethysmograms", in *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, 2017, pp. 56–62. DOI: 10.1109/FG.2017.16.
- [6] W.-Y. Yau, H.-T. Tran, E.-K. Teoh, and J.-G. Wang, "Fake finger detection by finger color change analysis", in *Advances in Biometrics: International Conference, ICB 2007, Seoul, Korea, August 27-29, 2007. Proceedings*, S.-W. Lee and S. Z. Li, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 888–896, ISBN: 978-3-540-74549-5. doi: 10.1007/978-3-540-74549-5_93. [Online]. Available: https://doi.org/10.1007/978-3-540-74549-5_93.
- [7] M. Drahansky, M. Dolezel, J. Vana, E. Brezinova, J. Yim, and K. Shim, "New optical methods for liveness detection on fingers", *BioMed research international*, vol. 2013, 2013.
- [8] E. Marasco and A. Ross, "A survey on antispoofing schemes for fingerprint recognition systems", *ACM Comput. Surv.*, vol. 47, no. 2, 28:1–28:36, Nov. 2014, ISSN: 0360-0300. DOI: 10.1145/2617756. [Online]. Available: <http://doi.acm.org/10.1145/2617756>.
- [9] K. K. Venkatasubramanian, A. Banerjee, and S. K. S. Gupta, "Plethysmogram-based secure inter-sensor communication in body area networks", in *MILCOM 2008 - 2008 IEEE Military Communications Conference*, 2008, pp. 1–7. DOI: 10.1109/MILCOM.2008.4753199.

- [10] C. C. Y. Poon, Y.-T. Zhang, and S.-D. Bao, “A novel biometrics method to secure wireless body area sensor networks for telemedicine and m-health”, *IEEE Communications Magazine*, vol. 44, no. 4, pp. 73–81, 2006, ISSN: 0163-6804. DOI: 10.1109/MCOM.2006.1632652.
- [11] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004, ISSN: 1051-8215. DOI: 10.1109/TCSVT.2003.818349.
- [12] R. McCraty, M. Atkinson, W. A. Tiller, G. Rein, and A. D. Watkins, “The effects of emotions on short-term power spectrum analysis of heart rate variability”, *The American Journal of Cardiology*, vol. 76, no. 14, pp. 1089–1093, 1995, ISSN: 0002-9149. DOI: [http://dx.doi.org/10.1016/S0002-9149\(99\)80309-9](http://dx.doi.org/10.1016/S0002-9149(99)80309-9). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0002914999803099>.
- [13] U. Yadav, S. N. Abbas, and D. Hatzinakos, “Evaluation of PPG biometrics for authentication in different states”, in *2018 International Conference on Biometrics (ICB)*, 2018, pp. 277–282. DOI: 10.1109/ICB2018.2018.00049.
- [14] A. B. Hertzman, “Observations on the finger volume pulse recorded photoelectrically”, *Am. J. Physiol.*, vol. 119, pp. 334–335, 1937.
- [15] J. Allen, “Photoplethysmography and its application in clinical physiological measurement”, *Physiol Meas*, vol. 28, no. 3, pp. 1–39, 2007.
- [16] J. Allen and A. Murray, “Variability of photoplethysmography peripheral pulse measurements at the ears, thumbs and toes”, *IEE Proceedings - Science, Measurement and Technology*, vol. 147, no. 6, pp. 403–407, 2000, ISSN: 1350-2344. DOI: 10.1049/ip-smt:20000846.
- [17] M. Elgendi, “On the analysis of fingertip photoplethysmogram signals”, *Curr Cardiol Rev*, vol. 8, no. 1, pp. 14–25, 2012.
- [18] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, “Multiparameter respiratory rate estimation from the photoplethysmogram”, *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946–1953, 2013, ISSN: 0018-9294. DOI: 10.1109/TBME.2013.2246160.
- [19] R. R. Anderson and J. A. Parrish, “The optics of human skin”, *Journal of Investigative Dermatology*, vol. 77, no. 1, pp. 13–19, 1981, ISSN: 0022-202X. DOI: <http://dx.doi.org/10.1111/1523-1747.ep12479191>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0022202X15461251>.
- [20] J. Lee, K. Matsumura, K. i. Yamakoshi, P. Rolfe, S. Tanaka, and T. Yamakoshi, “Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion”, in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 1724–1727. DOI: 10.1109/EMBC.2013.6609852.
- [21] V. Roberts, “Photoplethysmography- fundamental aspects of the optical properties of blood in motion”, *Transactions of the Institute of Measurement and Control*, vol. 4, no. 2, pp. 101–106, 1982. DOI: 10.1177/014233128200400205. eprint: <https://doi.org/10.1177/014233128200400205>. [Online]. Available: <https://doi.org/10.1177/014233128200400205>.

- [22] A. Kamal, J. Harness, G. Irving, and A. Mearns, “Skin photoplethysmography — a review”, *Computer Methods and Programs in Biomedicine*, vol. 28, no. 4, pp. 257–269, 1989, ISSN: 0169-2607. DOI: [http://dx.doi.org/10.1016/0169-2607\(89\)90159-4](http://dx.doi.org/10.1016/0169-2607(89)90159-4). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0169260789901594>.
- [23] U. Schultz-Ehrenburg and V. Blazek, “Value of quantitative photoplethysmography for functional vascular diagnostics. Current status and prospects”, *Skin Pharmacol. Appl. Skin Physiol.*, vol. 14, no. 5, pp. 316–323, 2001.
- [24] M. Huelsbusch and V. Blazek, “Contactless mapping of rhythmical phenomena in tissue perfusion using PPGI”, in *Proc. SPIE*, vol. 4683, 2002, pp. 110–117.
- [25] F. P. Wieringa, F. Mastik, and A. F.W.v. d. Steen, “Contactless multiple wavelength photoplethysmographic imaging: A first step toward “spo2 camera” technology”, *Annals of Biomedical Engineering*, vol. 33, no. 8, pp. 1034–1041, 2005, ISSN: 1573-9686. DOI: 10.1007/s10439-005-5763-2. [Online]. Available: <https://doi.org/10.1007/s10439-005-5763-2>.
- [26] Y. Y. Gu, Y. Zhang, and Y. T. Zhang, “A novel biometric approach in human verification by photoplethysmographic signals”, in *4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 2003.*, 2003, pp. 13–14. DOI: 10.1109/ITAB.2003.1222403.
- [27] Y. Y. Gu and Y. T. Zhang, “Photoplethysmographic authentication through fuzzy logic”, in *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, 2003.*, 2003, pp. 136–137. DOI: 10.1109/APBME.2003.1302621.
- [28] J. Yao, X. Sun, and Y. Wan, “A pilot study on using derivatives of photoplethysmographic signals as a biometric identifier”, in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 4576–4579. DOI: 10.1109/IEMBS.2007.4353358.
- [29] P. Shi, S. Hu, A. Echiadis, V. Azorin-Peris, J. Zheng, and Y. Zhu, “Development of a remote photoplethysmographic technique for human biometrics”, © SPIE, 2009.
- [30] A. Bonissi, R. D. Labati, L. Perico, R. Sassi, F. Scotti, and L. Sparagino, “A preliminary study on continuous authentication methods for photoplethysmographic biometrics”, in *2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, 2013, pp. 28–33. DOI: 10.1109/BIOMS.2013.6656145.
- [31] A. R. Kavsaoglu, K. Polat, and M. R. Bozkurt, “A novel feature ranking algorithm for biometric recognition with PPG signals”, *Computers in Biology and Medicine*, vol. 49, pp. 1–14, 2014, ISSN: 0010-4825. DOI: <http://dx.doi.org/10.1016/j.compbiomed.2014.03.005>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010482514000687>.
- [32] A. Lee and Y. Kim, “Photoplethysmography as a form of biometric authentication”, in *2015 IEEE SENSORS*, 2015, pp. 1–2. DOI: 10.1109/ICSENS.2015.7370629.
- [33] S. P. M. Namini and S. Rashidi, “Implementation of artificial features in improvement of biometrics based PPG”, in *2016 6th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2016, pp. 342–346. DOI: 10.1109/ICCKE.2016.7802164.

- [34] N. I. M. Nadzri, K. A. Sidek, and R. M. Nor, “Biometric identification for twins using photoplethysmogram signals”, in *2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*, 2016, pp. 320–324. doi: 10.1109/ICT4M.2016.071.
- [35] S. Chakraborty and S. Pal, “Photoplethysmogram signal based biometric recognition using linear discriminant classifier”, in *2016 2nd International Conference on Control, Instrumentation, Energy Communication (CIEC)*, 2016, pp. 183–187. doi: 10.1109/CIEC.2016.7513792.
- [36] V. Jindal, J. Birjandtalab, M. B. Pouyan, and M. Nourani, “An adaptive deep learning approach for PPG-based identification”, in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 6401–6404. doi: 10.1109/EMBC.2016.7592193.
- [37] A. Sarkar, A. L. Abbott, and Z. Doerzaph, “Biometric authentication using photoplethysmography signals”, in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2016, pp. 1–7. doi: 10.1109/BTAS.2016.7791193.
- [38] T. Ohtsuki and H. Kamoi, “Biometric authentication using hand movement information from wrist-worn PPG sensors”, in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016, pp. 1–5. doi: 10.1109/PIMRC.2016.7794969.
- [39] P. Spachos, J. Gao, and D. Hatzinakos, “Feasibility study of photoplethysmographic signals for biometric identification”, in *2011 17th International Conference on Digital Signal Processing (DSP)*, 2011, pp. 1–5. doi: 10.1109/ICDSP.2011.6004938.
- [40] N. S.G. R. Salanke, N. Maheswari, A. Samraj, and S. Sadhasivam, “Enhancement in the design of biometric identification system based on photoplethysmography data”, in *2013 International Conference on Green High Performance Computing (ICGHPC)*, 2013, pp. 1–6. doi: 10.1109/ICGHPC.2013.6533909.
- [41] N. S. Girish Rao Salanke, N. Maheswari, and A. Samraj, “An enhanced intrinsic biometric in identifying people by photoplethysmography signal”, in *Proceedings of the Fourth International Conference on Signal and Image Processing 2012 (ICSIP 2012): Volume 1*, M. S and S. S. Kumar, Eds. India: Springer India, 2013, pp. 291–299, ISBN: 978-81-322-0997-3. doi: 10.1007/978-81-322-0997-3_27. [Online]. Available: https://doi.org/10.1007/978-81-322-0997-3_27.
- [42] N. A. L. Jaafar, K. A. Sidek, and S. N.A. M. Azam, “Acceleration plethysmogram based biometric identification”, in *2015 International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, 2015, pp. 16–21. doi: 10.1109/ICBAPS.2015.7292210.
- [43] S. N.A. M. Azam and K. A. Sidek, “Time variability analysis of photoplethysmogram biometric identification system”, *Indian Journal of Science and Technology*, vol. 9, no. 28, 2016, ISSN: 0974 -5645. [Online]. Available: <http://www.indjst.org/index.php/indjst/article/view/97731>.
- [44] J. da Silva Dias, I. Traore, V. G. Ferreira, and J. David, “Exploratory use of PPG signal in continuous authentication”, in *The Brazilian Symposium on Information and Computational Systems Security*, 2015.
- [45] T. Choudhary and M. S. Manikandan, “Robust photoplethysmographic (PPG) based biometric authentication for wireless body area networks and m-health applications”, in *2016 Twenty Second National Conference on Communication (NCC)*, 2016, pp. 1–6. doi: 10.1109/NCC.2016.7561152.

- [46] N. Karimian, M. Tehranipoor, and D. Forte, “Non-fiducial PPG-based authentication for healthcare application”, in *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2017, pp. 429–432. doi: 10.1109/BHI.2017.7897297.
- [47] N. Karimian, Z. Guo, M. Tehranipoor, and D. Forte, “Human recognition from photoplethysmography (PPG) based on non-fiducial features”, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 4636–4640. doi: 10.1109/ICASSP.2017.7953035.
- [48] J. Yathav, A. Bailur, A. K. Goyal, and Abhinav, “miBEAT based continuous and robust biometric identification system for on-the-go applications”, in *Proceedings of International Conference on Communication and Networks: ComNet 2016*, N. Modi, P. Verma, and B. Trivedi, Eds. Singapore: Springer Singapore, 2017, pp. 269–275, ISBN: 978-981-10-2750-5. doi: 10.1007/978-981-10-2750-5_28. [Online]. Available: https://doi.org/10.1007/978-981-10-2750-5_28.
- [49] J. Sancho, A. Alesanco, and J. Garca, “Photoplethysmographic authentication in long-term scenarios: A preliminary assessment”, in *EMBEC & NBC 2017: Joint Conference of the European Medical and Biological Engineering Conference (EMBEC) and the Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC), Tampere, Finland, June 2017*, H. Eskola, O. Väisänen, J. Viik, and J. Hyttinen, Eds. Singapore: Springer Singapore, 2018, pp. 1085–1088, ISBN: 978-981-10-5122-7. doi: 10.1007/978-981-10-5122-7_271. [Online]. Available: https://doi.org/10.1007/978-981-10-5122-7_271.
- [50] *BioSec.Lab databases*, <http://www.comm.utoronto.ca/~biometrics/databases.html>, Last accessed on 14 Sept 2017.
- [51] *Pulse sensor*, <https://store.plux.info/bitalino-sensors/42-pulsesensor.html>, Last accessed on 14 Sept 2017.
- [52] *Bitalino CORE BT*, <https://store.plux.info/bitalino-components/24-bitalino-revolution-core-mcubtpwr-810121705.html>, Last accessed on 14 Sept 2017.
- [53] *OpenSignals software*, <http://biosignalsplus.com/en/software>, Last accessed on 14 Sept 2017.
- [54] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “DEAP: A database for emotion analysis ;using physiological signals”, *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012, ISSN: 1949-3045. doi: 10.1109/T-AFFC.2011.15.
- [55] J. Pan and W. J. Tompkins, “A real-time QRS detection algorithm”, *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985, ISSN: 0018-9294. doi: 10.1109/TBME.1985.325532.
- [56] X. He, R. A. Goubran, and X. P. Liu, “Secondary peak detection of PPG signal for continuous cuffless arterial blood pressure measurement”, *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 6, pp. 1431–1439, 2014, ISSN: 0018-9456. doi: 10.1109/TIM.2014.2299524.
- [57] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans, “Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions”, *PLOS ONE*, vol. 8, no. 10, pp. 1–11, Oct. 2013. doi: 10.1371/journal.pone.0076585. [Online]. Available: <https://doi.org/10.1371/journal.pone.0076585>.

- [58] H. S. Shin, C. Lee, and M. Lee, “Adaptive threshold method for the peak detection of photoplethysmographic waveform”, *Computers in Biology and Medicine*, vol. 39, no. 12, pp. 1145 –1152, 2009, ISSN: 0010-4825. DOI: <https://doi.org/10.1016/j.combiomed.2009.10.006>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010482509001826>.
- [59] D. McDuff, S. Gontarek, and R. W. Picard, “Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera”, *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 12, pp. 2948–2954, 2014, ISSN: 0018-9294. DOI: 10.1109/TBME.2014.2340991.
- [60] G. Heusch, A. Anjos, and S. Marcel, “A reproducible study on remote heart rate measurement”, *CoRR*, vol. abs/1709.00962, 2017. arXiv: 1709.00962. [Online]. Available: <http://arxiv.org/abs/1709.00962>.
- [61] W. Louis and D. Hatzinakos, “Enhanced binary patterns for electrocardiogram (ECG) biometrics”, in *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2016, pp. 1–4. DOI: 10.1109/CCECE.2016.7726725.
- [62] W. Louis, M. Komeili, and D. Hatzinakos, “Continuous authentication using one-dimensional multi-resolution local binary patterns (1DMRLBP) in ECG biometrics”, *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2818–2832, 2016, ISSN: 1556-6013. DOI: 10.1109/TIFS.2016.2599270.
- [63] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, “On evaluating ECG biometric systems: Session-dependence and body posture”, *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 2002–2013, 2014, ISSN: 1556-6013. DOI: 10.1109/TIFS.2014.2360430.
- [64] F. Agrafioti and D. Hatzinakos, “ECG based recognition using second order statistics”, in *6th Annual Communication Networks and Services Research Conference (CNSR 2008)*, 2008, pp. 82–87. DOI: 10.1109/CNSR.2008.38.
- [65] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, “An introduction to kernel-based learning algorithms”, *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 181–201, 2001, ISSN: 1045-9227. DOI: 10.1109/72.914517.
- [66] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997, ISSN: 0162-8828. DOI: 10.1109/34.598228.
- [67] W. Zhao, R. Chellappa, and A. Krishnaswamy, “Discriminant analysis of principal components for face recognition”, in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 336–341. DOI: 10.1109/AFGR.1998.670971.
- [68] H. Yu and J. Yang, “A direct LDA algorithm for high-dimensional data — with application to face recognition”, *Pattern Recognition*, vol. 34, no. 10, pp. 2067 –2070, 2001, ISSN: 0031-3203. DOI: [https://doi.org/10.1016/S0031-3203\(00\)00162-X](https://doi.org/10.1016/S0031-3203(00)00162-X). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S003132030000162X>.
- [69] M. Turk and A. Pentland, “Eigenfaces for recognition”, *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

- [70] J. Yang, Z. Jin, J. yu Yang, D. Zhang, and A. F. Frangi, “Essence of kernel fisher discriminant: KPCA plus LDA”, *Pattern Recognition*, vol. 37, no. 10, pp. 2097 –2100, 2004, ISSN: 0031-3203. doi: <https://doi.org/10.1016/j.patcog.2003.10.015>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320303004205>.
- [71] G. Baudat and F. Anouar, “Generalized discriminant analysis using a kernel approach”, *Neural computation*, vol. 12, no. 10, pp. 2385–2404, 2000.
- [72] J. Lu, K. Plataniotis, and A. Venetsanopoulos, “Kernel discriminant learning with application to face recognition”, in *Support Vector Machines: Theory and Applications*, L. Wang, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 275–296, ISBN: 978-3-540-32384-6. doi: 10.1007/10984697_13. [Online]. Available: https://doi.org/10.1007/10984697_13.
- [73] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, “Face recognition using kernel direct discriminant analysis algorithms”, *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 117–126, 2003, ISSN: 1045-9227. doi: 10.1109/TNN.2002.806629.
- [74] J. Lu, K. Plataniotis, and A. Venetsanopoulos, “Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition”, *Pattern Recognition Letters*, vol. 26, no. 2, pp. 181 –191, 2005, ISSN: 0167-8655. doi: <https://doi.org/10.1016/j.patrec.2004.09.014>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865504002260>.
- [75] I. Odinaka, P. H. Lai, A. D. Kaplan, J. A. O’Sullivan, E. J. Sirevaag, S. D. Kristjansson, A. K. Sheffield, and J. W. Rohrbaugh, “ECG biometrics: A robust short-time frequency analysis”, in *2010 IEEE International Workshop on Information Forensics and Security*, 2010, pp. 1–6. doi: 10.1109/WIFS.2010.5711466.
- [76] D. M. J. Tax, “One-class classification”, 2001.
- [77] C. J. Burges, “A tutorial on support vector machines for pattern recognition”, *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998, ISSN: 1573-756X. doi: 10.1023/A:1009715923555. [Online]. Available: <https://doi.org/10.1023/A:1009715923555>.
- [78] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006, ISBN: 0387310738.
- [79] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1.
- [80] A. Ben-Hur and J. Weston, “A user’s guide to support vector machines”, in *Data Mining Techniques for the Life Sciences*, O. Carugo and F. Eisenhaber, Eds. Totowa, NJ: Humana Press, 2010, pp. 223–239, ISBN: 978-1-60327-241-4. doi: 10.1007/978-1-60327-241-4_13. [Online]. Available: https://doi.org/10.1007/978-1-60327-241-4_13.
- [81] H. He and E. A. Garcia, “Learning from imbalanced data”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009, ISSN: 1041-4347. doi: 10.1109/TKDE.2008.239.
- [82] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique”, *J. Artif. Int. Res.*, vol. 16, no. 1, pp. 321–357, Jun. 2002, ISSN: 1076-9757. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1622407.1622416>.

- [83] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, “Support vector method for novelty detection”, in *Proceedings of the 12th International Conference on Neural Information Processing Systems*, ser. NIPS’99, Denver, CO: MIT Press, 1999, pp. 582–588. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3009657.3009740>.
- [84] D. M. Tax and R. P. Duin, “Support vector data description”, *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004, ISSN: 1573-0565. doi: 10.1023/B:MACH.0000008084.60811.49. [Online]. Available: <https://doi.org/10.1023/B:MACH.0000008084.60811.49>.
- [85] A. Criminisi, E. Konukoglu, and J. Shotton, “Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning”, Tech. Rep., 2011. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/decision-forests-for-classification-regression-density-estimation-manifold-learning-and-semi-supervised-learning/>.
- [86] C. Leistner, A. Saffari, J. Santner, and H. Bischof, “Semi-supervised random forests”, in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 506–513. doi: 10.1109/ICCV.2009.5459198.
- [87] L. Breiman, “Random forests”, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, ISSN: 1573-0565. doi: 10.1023/A:1010933404324. [Online]. Available: <https://doi.org/10.1023/A:1010933404324>.
- [88] ——, *Classification and regression trees*. Routledge, 2017.
- [89] J. Wang, H. Lu, K. Plataniotis, and J. Lu, “Gaussian kernel optimization for pattern classification”, *Pattern Recognition*, vol. 42, no. 7, pp. 1237 –1247, 2009, ISSN: 0031-3203. doi: <https://doi.org/10.1016/j.patcog.2008.11.024>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320308005013>.
- [90] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, “Learning the kernel matrix with semidefinite programming”, *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, Dec. 2004, ISSN: 1532-4435. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1005332.1005334>.
- [91] L. Bottou and C. jen Lin, *Support vector machine solvers*, 2006.
- [92] K. N. Plataniotis, D. Hatzinakos, and J. K. M. Lee, “ECG biometric recognition without fiducial detection”, in *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, 2006, pp. 1–6. doi: 10.1109/BCC.2006.4341628.