Title
A Study into the Feasibility of Biometric Identification Through ECG

Author
Cordes, Kyle

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

A STUDY INTO THE FEASABILITY OF BIOMETRIC IDENTIFICATION THROUGH ECG

A thesis submitted in partial satisfaction of the requirements for the degree of

Master of Science

in

COMPUTER ENGINEERING

by

Kyle Jeremy Cordes

September 2018

The thesis of Kyle Jeremy Cordes is approved:

Prof. Matthew Guthaus, Chair

Prof. Patrick Mantey

Dr. Ali Adabi

Lori Kletzer
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

A STUDY INTO THE FEASABILITY OF BIOMETRIC IDENTIFICATION THROUGH ECG

by

Kyle Jeremy Cordes

This research proposes the idea that an electrocardiogram (ECG) is unique among individuals. A group of 84 individuals was considered and classified using two classification techniques, Linear Discriminant Analysis and a Feed Forward Neural Network. These classifiers are used to identify uniqueness in an individual's ECG in both the time and frequency domain. This research found that we can classify this set of data with a 93% accuracy.

To my family and friends.

Acknowledgments

The work done throughout this research would not have been possible without all of the support that was provided by everyone around me.

I would like to start by first extending my deepest gratitude to my advisor, Professor Matthew Guthaus, who provided me with encouragement, support and direction throughout my degree. Next, I would like to thank Professor Patrick Mantey for being part of the reading committee and giving feedback to help me further the development of this thesis. I would also like to thank Doctor Ali Adabi for being part of the reading committee and providing his expertise in the subject during the writing process.

Lastly, I would like to thank my loving family and friends for their support. In particular, my parents Kim and Joel, who have been nothing but supportive throughout my entire life and academic career. In addition to this I would like to thank my brother Sam, for always being a phone call away whenever I needed it. I would also like to thank my girlfriend Andi for all the support, love and edits she has provided along the way.

All of these people have been an integral part of my success and I cannot ever thank them enough.

Chapter 1

Introduction

With the increasing availability of accurate and mobile sensing devices, we are seeing a wide variety of applications. One of these applications comes in the form of security. An electrocardiogram, known commonly as ECG or EKG, is a measurement of the electrical impulses in the heart as they move from the atrium to the ventricle [14]. Due to the inherent uniqueness of every human, it may be possible to use ECG to enhance already existing biometric security technology to further identify an individual.

An ECG consists of three main segments or waves as shown in Figure 1.1. These are the P-wave, QRS complex, and the T-wave [14]. There is also a U wave but this is not always observable [14]. The P-wave, in normal heart function, is a smooth upwards peak that is caused by atrial depolarization. The QRS complex is generally a small downwards spike, followed by a large upwards spike, which is then followed by another small downwards spike. This complex represents ventricular depolarization and contraction. Next, is a smoother upward wave known as the T-wave, that represents ventricular re-polarization. This is lastly followed by the U wave, which represents the recovery of the Purkinje conduction fibers, but is often unobservable [14]. These waves can be seen graphically in Figure 1.1.

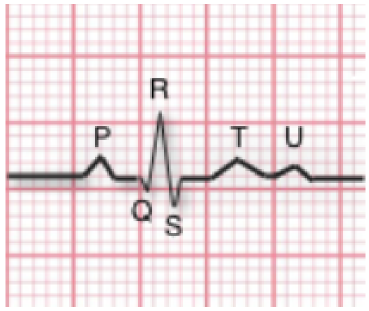


Figure 1.1: An ECG signal consists of a P-wave, QRS complex, T, and U waves. These represent electrical signals produced as the heart beats.

A regular or normal heartbeat exists if there are no arrhythmias or other irregularities, which can cause inconsistent heartbeat signatures. For this reason, irregular heartbeats are either not used or removed from the study.

The idea that an ECG signal is unique to the individual implies that certain features about the signal are repeatable across multiple heartbeats. An example of these repeatable features is seen in Figure 1.2. This figure was generated by splitting up an individuals ECG signal into segments that contain the PQRST fragments. These segments are then centered about the R-peak and a heat map is generated where a redder signal represents a higher incidence of a given value. The more blue the signal, the less occurrences of that value exist. This visualization of an individual's ECG gives us an understanding that their ECG signature can remain relatively consistent between beats, and therefore has the possibility of being unique.

Previous work in the area of identifying uniqueness in the ECG signal is documented in Tatiana Lugovaya's thesis, "Biometric Human Identification Based On ECG" [13]. In Lugovaya's thesis, Linear Discriminant Analysis was used on a reduced feature-space in the time domain to classify individuals. The research discussed in the following chapters builds upon Lugovaya's findings and accomplishes several goals to further the field of ECG identification research. The first

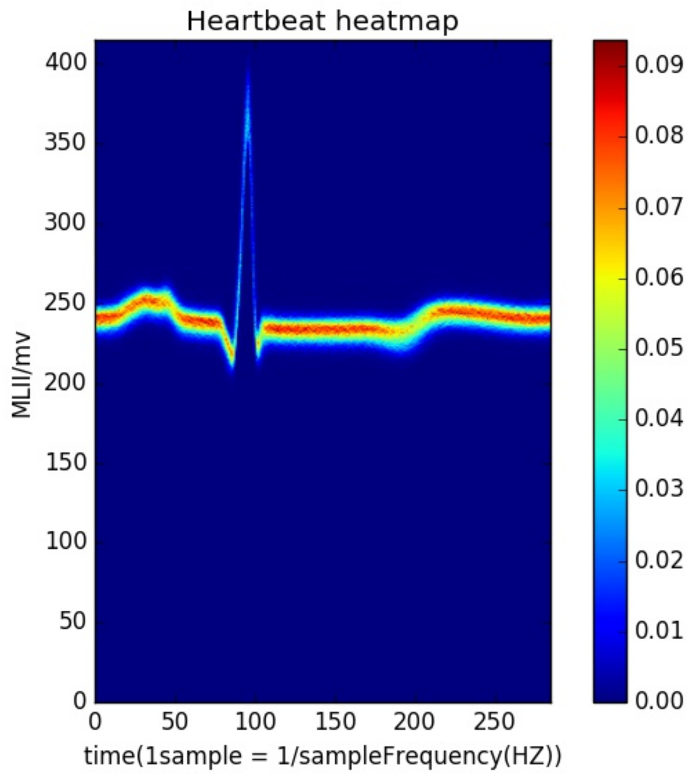


Figure 1.2: This heat map of multiple PQRSTU segments from a single individual shows that the ECG signal has high regularity in the P and T waves, as well as the segments connecting them when centered about its R-peak.

is verifying the work performed by Lugovaya to provide a baseline for the other classification techniques used. The second is to try several different classification techniques with different feature-space formations and compare those to the baseline set by Lugovaya. Lastly, the frequency domain in both a reduced and unreduced feature space is used to determine the effects of domain type on the classification accuracy.

Chapter 3 of this research focuses on the formation of the feature space. With most machine learning algorithms, the parameters are trained by using many pieces of the type of data that is desired to be classified. When using algorithms that use this technique, it is important to separate the data to avoid testing the classifier on the data it was trained on. It is for this reason that a training and test set are created. This means that a certain percentage (approximately 70%) of the data is used for training, and then another percentage (approximately 30%) for testing. The precise break down of these designated records can be see in Chapter 3: Feature Space Formation. Some classification algorithms can see improvement when the feature space, or the number of identifying characteristics, are reduced. This research looks into classifying on both a reduced and unreduced feature space. Although there are many ways to reduce a feature space, for this purpose, Principle Component Analysis (PCA) was the reduction technique investigated.

Lastly, this research delves into two different classification techniques. These are Linear Discriminant Analysis (LDA) and Feed Forward Neural Networks (FFNN). Linear Discriminant Analysis is used as a baseline metric to see how changes in the feature space and use of a different classifier effect the overall accuracy of classification. Both of the classifiers were trained on a reduced and unreduced feature space in the time and frequency domains.

Chapter 2

Data Preprocessing

This chapter describes how the ECG signals are measured and how they are taken from raw unprocessed signals to usable ECG segments that are used to form the feature space for classification. It is important to filter or process the signals because there are many elements of noise that can enter the system. Some of these sources of noise can include power line noise and movement in the patient at the time of recording. This noise can be removed with basic signal processing filter techniques such as low, high, and bandpass filters. The data also needs to be normalized because while recording, the amplitude of the signals can vary across multiple recordings of an individual patient due to the connections that the leads have. For this reason, amplitude is less important than wave shape. In turn, a normalized measurement standard is created across multiple individuals and recordings. This normalization enhances the differences in the unique characteristics of each ECG signature instead of differences in the recording environment.

To start, the ECG signal is analyzed in the frequency domain to determine the type of noise that is present, in order to help better design the cutoff frequencies for the filters that are used. A common noise source that can be seen is visualized as abrupt changes in the ECG signal. These noise sources are caused by disturbances during recording and can be viewed as outliers and removed accordingly. Next, the segment between the Q-peak and T-wave changes proportionally with the R-peak to R-peak (RR) duration or heart rate [14]. This research aims to correct for this in order to gain independence from heart rate, which creates an environment in which an individuals heart rate does not effect its classification. To accomplish this, two formulas are considered, Framinham's and Bazette's formulas [8][11]. On a parallel path, the frequency domain of the signal is considered because this research aims to classify the signal using both the time and frequency domains of the signal. In order to get an accurate FFT of the signal a rectangular window of the heartbeats is used. Lastly, in order to normalize the segments the unit norm of all of the vectors are taken. After these techniques are applied, the segments of the ECG are in the range $[-1, 1]$.

2.1 Benchmark Overview

This chapter covers how the data was processed before it is classified using the various classification techniques discussed in a later chapter. Before processing can happen the data is recorded or taken from a database considering multiple factors such as measurement type and whether or not the heartbeat is irregular.

In this research, we take from a database posted on physio-bank for the purposes of ECG Identification [3]. This data contains 310 ECG recordings obtained from 90 people. Each of the recordings is taken from ECG lead I for an average of 20 seconds. The signal was recorded with a sample rate of 500 Hz with a 12 bit resolution and a nominal $\pm 10mV$ range. The data was collected from 44 men and 46 woman in the age range of 13 to 75. However, due to little amount of data being available for some of the individuals, some of the people were removed leaving the data set with 84 total individuals. A record for each person was collected over a period of 6 months. Lead I is the potential difference between the right and left hands [13]. This database was chosen because it provides a database that spans multiple individuals across multiple age groups and sexes.

2.2 Filtering Signal

First, a recording needs to be filtered to remove unwanted frequencies introduced through several means. In order to filter this signal a region of interest needs to be selected to tune the bandwidth of the chosen filter. A bandpass filter is used in order to remove both low and high frequency noise. In normal heart function, a normal band of frequency to see a heartbeat between is 60 to 100 beats per minute (BPM) [10]. However, when recording a signal using the first lead, there is noise introduced from the power supply. In addition, there is noise introduced from the patient's breathing and movement. In order to remove this noise a filter needs to be tuned to suppress these signals but also to keep the gain of the ECG minimally changed as too much information could be lost.

To chose the bandwidth of the bandpass filter, the noise was first examined in the frequency domain performing the Fast-Fourier transform or FFT. This can be seen in Figure 2.1. To interpret this it is important to remember what was mentioned above, that a normal heart beats between 60 to 100 BPM. This relates to a frequency of 1 to 1.33 Hz. Looking at the region of interest in Figure 2.1, a filter needs to be designed to keep this region relatively unaltered while filtering out the power supply noise and DC offset. The power supply noise can be seen at $50Hz$. For this reason, a filter with the bands $0.6Hz$ and $20Hz$ was chosen. After this filter was applied, the signal has significantly less noise while still maintaining its defining characteristics. In Figure 2.2, it can be seen that DC signals get filtered out by the designed bandpass filter. The high frequency noise is also filtered out

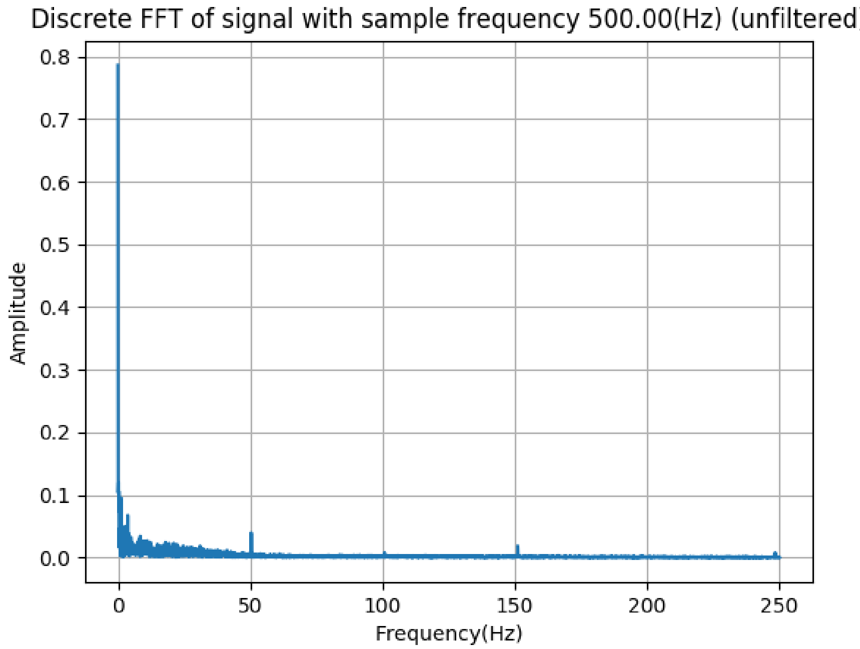


Figure 2.1: Unfiltered 20 second ECG signal with noise frequency spikes seen at 50, 150, and 250 Hz

by the bandpass filter, see Figure 2.3. In the frequency domain the noise spikes can be observed as removed in Figure 2.4

2.3 Extracting Segments and Removing Outliers

In this research, the ECG signals are classified on a segment basis; that is, each beat or PQRSTU segment is classified individually, then a majority vote of these segments is used to classify an individual. In order to do this, the segments need to be extracted in a way such that all the unique characteristics are still present in the segment. This research focuses on using all of the defining features of the PQRSTU segment (P-wave, QRS complex and a T-wave) to classify a person. Thus it is important that each segment must include each of these pieces. Although filtering takes care of much of the noise within the signal there are still problems that could not be removed with the bandpass filter. The first of these is

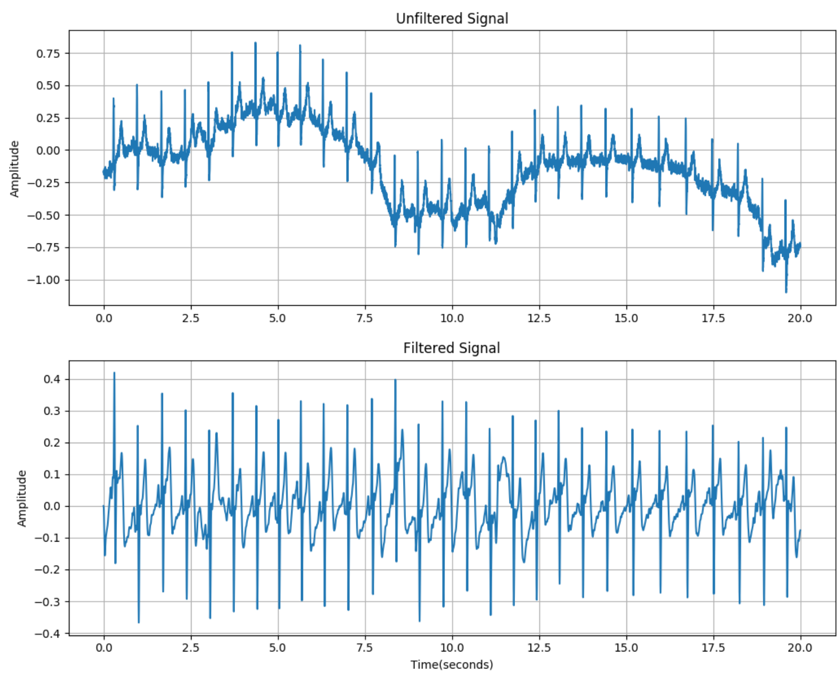


Figure 2.2: Filtered (low cut off = $0.6Hz$, high cut off = $20Hz$) and Unfiltered 20 second ECG signal. The unfiltered signal (top) contains a lower frequency noise bias that is removed by the bandpass filter producing a smoother filtered signal (bottom).

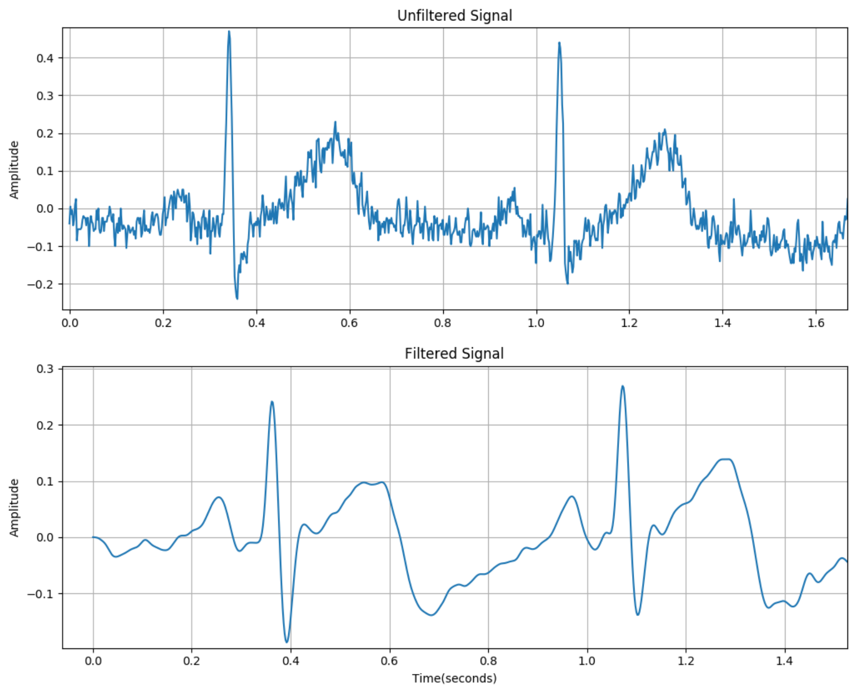


Figure 2.3: Filtered (low cut off $= 0.6Hz$, high cut off $= 20Hz$) and Unfiltered ECG signal containing two beats of the heart. The unfiltered signal (top), contains high frequency noise that is unwanted. The filtered signal (bottom) shows that the band pass filter has removed this high frequency noise, significantly smoothing the signal.

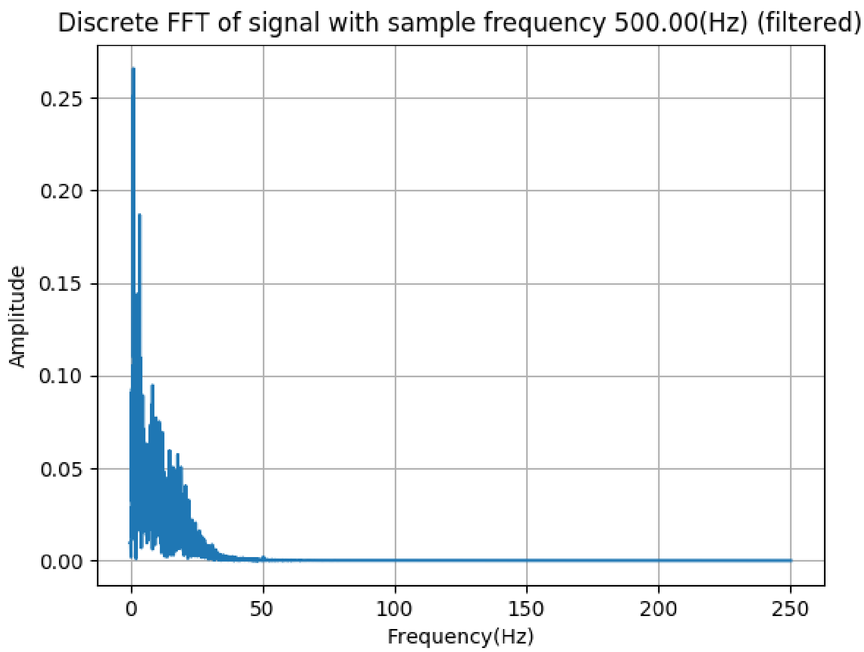


Figure 2.4: Filtered(low cut off = $0.6Hz$, high cut off = $20Hz$) FFT of ECG signal. This shows that the frequency noise spikes seen in the previous unfiltered FFT transformed signal are removed by the band pass filter. The low frequency near DC frequency spikes have also been greatly reduced.

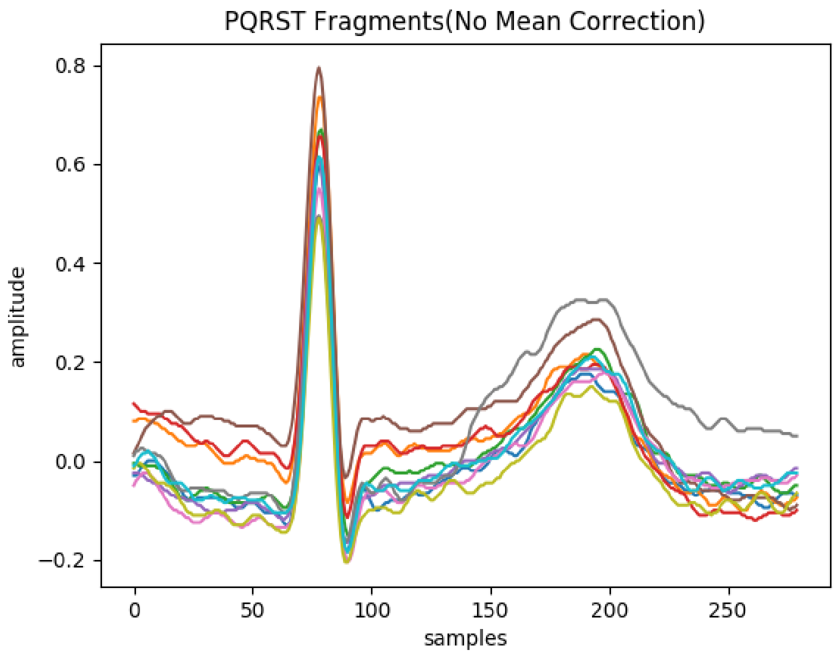


Figure 2.5: When no mean correction has been applied, baseline drift occurs causing the fragments to be offset from one another.

baseline drift. Baseline drift occurs due to respiration or other body movements as well as errors within the recording equipment itself. There are also discrete changes in the signal such as movement from the patient at time of recording that cause certain segments to be drastically different than the others. These are considered outliers and are removed from the data set.

To start, the segments were extracted and placed on top of each other for visualization purposes. As seen in Figure 2.5, the segments are seen to have the same duration between the peaks and valleys of the ECG signal, but differ in the their baseline. To fix this, a mean segment is constructed and used to remove the baseline drift. Once the mean segment is calculated the baseline drift is corrected by subtracting off the mean from each individual signal. The result of this manipulation can be seen in Figure 2.6.

After the the segments are extracted and the baseline drift is corrected, the

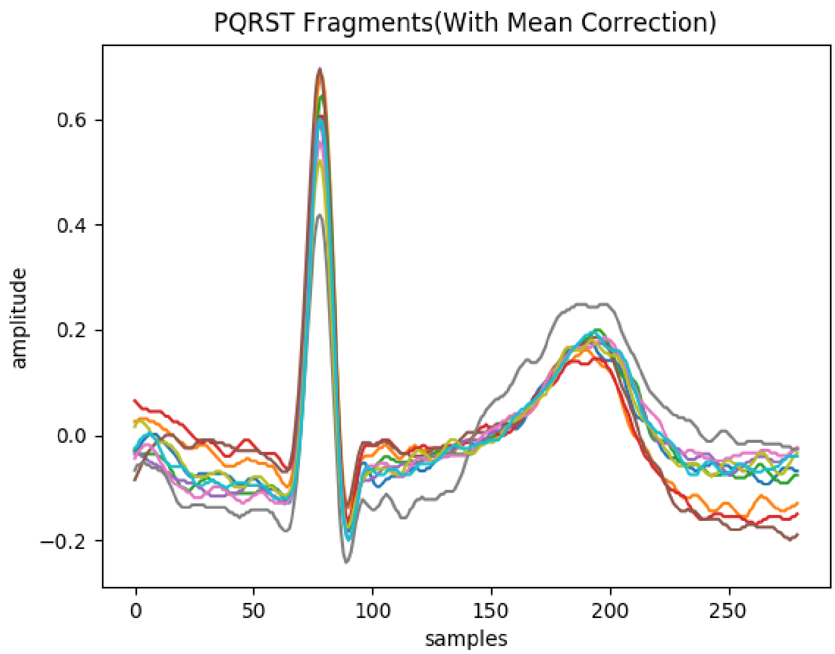


Figure 2.6: When mean correction is applied to the PQRST fragments, it can be seen that the fragments are no longer offset from each other. However, as seen in this figure, there are still some segments that are not uniform with the rest. These segments are considered outliers.

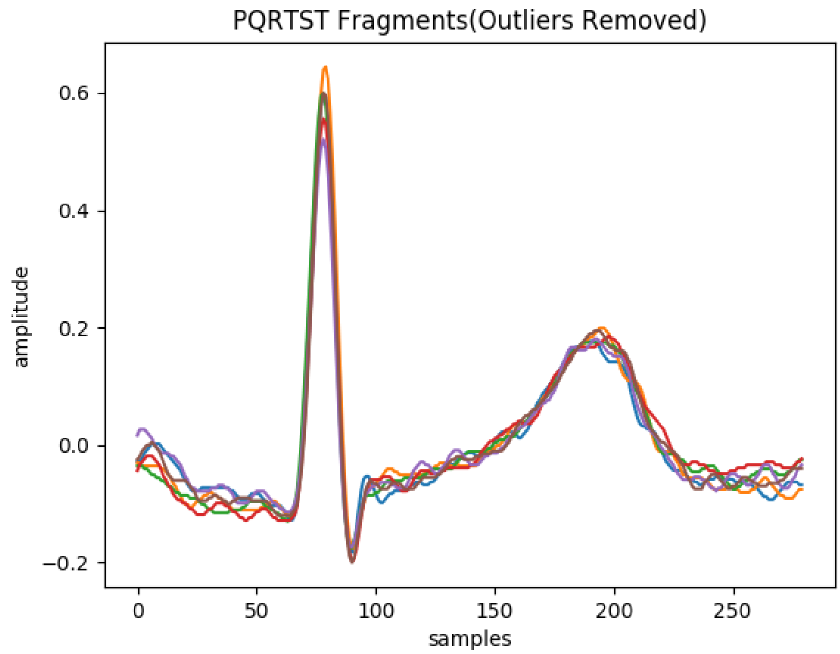


Figure 2.7: This figure shows that after the outliers have been removed the fragments are close to identical in general shape.

outlier segments need to be removed. In Figure 2.6, there are some segments that can still be seen diverging from the mean segment. In order to remove them, another mean segment is created and for each segment a magnitude is created that represents the distance away from the mean segment. Then segments whose magnitude are within a certain threshold magnitude are kept while those that are above this certain threshold are removed. This threshold is then tuned until a desired result is met. This culling of outlier segments can be seen in Figure 2.7, in which some of the outlier segments seen in Figure 2.6 have been removed leaving signals that represent the heart's characteristics rather than the recording equipment's characteristics.

2.4 Q-T Fragment Normalization

One of the largest variabilities in the ECG signal is the heart rate. When the heart rate changes, the durations between certain portions of the ECG segments will change. Obviously the RR interval changes, but since we are classifying based on individual PQRSTU fragments, the variation in the RR interval is not part of the feature-space. Studies have shown that there is little to no variation in the time interval between the PQRSTU segments of the ECG. However, the Q-T interval has been shown to vary based on heart rate [14]. The QT interval is inversely proportional to the heart rate, meaning that as heart rate increases the QT interval duration decreases. As heart rate decreases, the QT interval increases.

There are several methods that clinicians use to normalize the QT interval, and for this research Framingham and Bazett's formulas were both considered [8][11]. These formulas were used to calculate a normalized version of the QT interval at which point we adjust and re-sample the interval to create time-normalized PQRSTU segments of the individual's ECG signal. The formulas are dependent on the current duration of the RR interval (heart rate) as well as the current QT interval duration. A new interval is then calculated and the signal is adjusted accordingly. As seen in the table 2.1, varying heart rates and QT intervals output similar QT intervals after correction. These intervals were calculated using Bazzett's formula,

$$QT_{Bazett} = \frac{QT}{\sqrt{RR}} \tag{2.1}$$

as well as Framinham's formula,

$$QT_{Framingham} = QT + 0.154 * (1 - RR) \tag{2.2}$$

Framingham	Signals		
Heartrate(BPM)	68.807	94.637	57.692
QT interval original(second	0.244	0.206	0.264
QT interval corrected(seco	0.252	0.256	0.258

Bazette	Signals		
Heartrate(BPM)	68.807	94.637	57.692
QT interval original(second	0.244	0.206	0.264
QT interval corrected(seco	0.252	0.257	0.259

Table 2.1: Corrected QT time intervals based on heartrate using both Framingham and Bazett formula.

After the QT interval has been recalculated, the segments need to be reformed with this new interval. In order to accomplish this a re-sampling algorithm is used to create the new QT segment and attach it to the existing segment. This re-sampled signal is either used to expand or contract the existing QT-interval. The reason that the adjusted ECG signal needs to be re-sampled is that it is important to maintain a uniform sampling rate between each point in the signal. Then when the feature space is formed we are time independent and instead focus on the number of features in each segment. After re-sampling, the duration between each point is equivalent and the feature space can be formed to have n-features or can be reduced using feature reduction techniques discussed later. Once the wave is re-sampled the signal's tailing edge is cut to maintain a consistent 250 samples across all segments. The sample size was chosen because it encompasses all of the unique characterizing segments of an ECG segment, while not including characteristics from either the preceding or trailing PQRSTU fragments. The PQRSTU segments before correction can be seen in Figure 2.8 and the segments after correction using Framingham's formula can be seen in Figure 2.10, while using Bazett's formula can be seen in Figure 2.9. As seen in these figures the QT segments vary in length, but after the correction is applied they are roughly

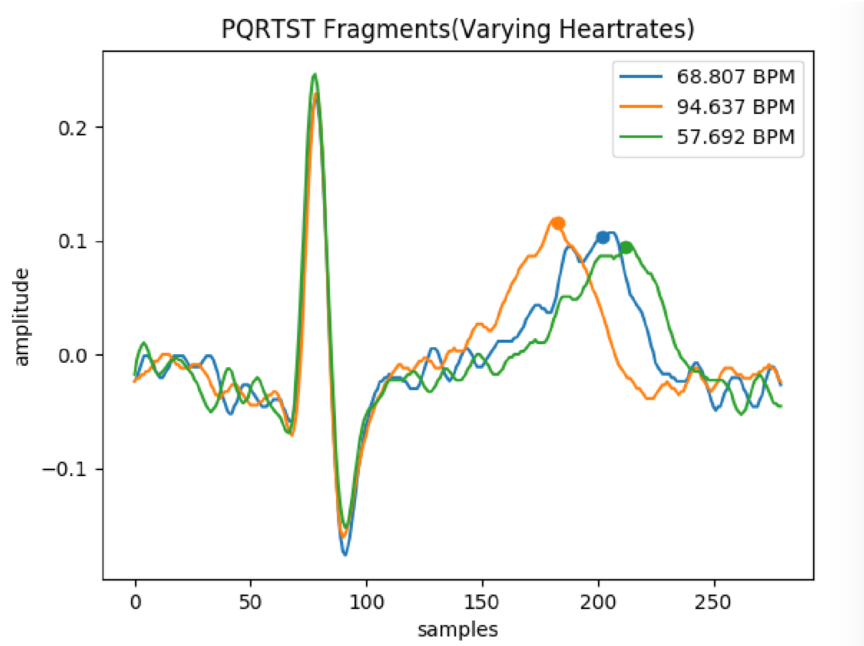


Figure 2.8: This figure shows that a varying heart rate changes the duration or length of the QT interval. This needs to be corrected in order to create heart rate independent PQRST fragments

equivalent durations.

2.5 Frequency Data Extraction

In addition to analyzing in the time domain, this research set out to determine if there are any improvements in classification accuracy to be gained by forming a feature space in the frequency domain. There are several challenges to overcome when moving the ECG signals into the time domain. As mentioned earlier in this chapter, the signals are chopped up into individual PQRSTU segments and then the QT segments are corrected. This would not be an issue if the Fast Fourier Transform (FFT) was only performed on a single segment. However, due to properties of the FFT and aliasing occurring by taking the FFT, multiple segments in succession need to be taken in order to obtain a more accurate frequency domain

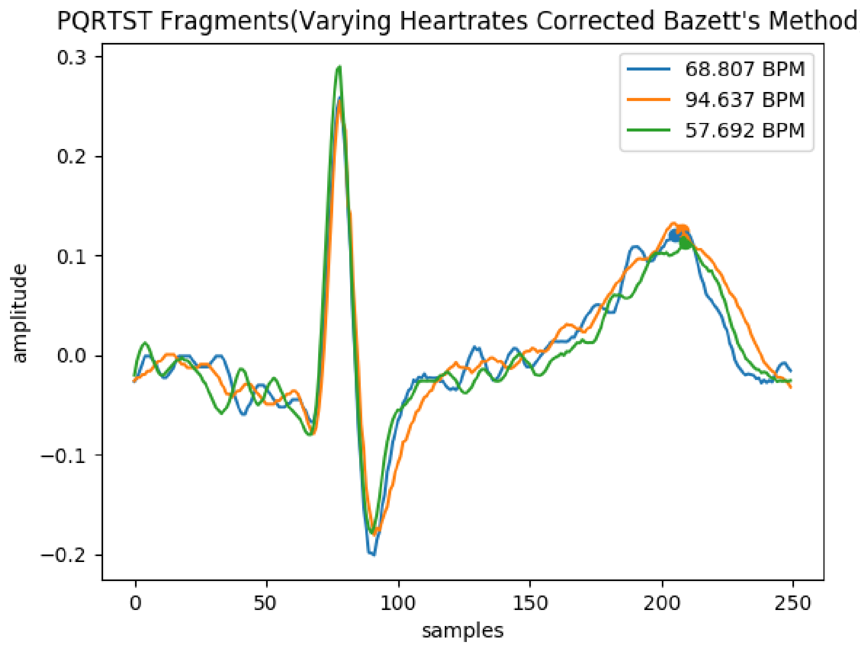


Figure 2.9: After applying Bazett's formula and re-sampling, the corrected QT segment is shown to be normalized regardless of the heart rate.

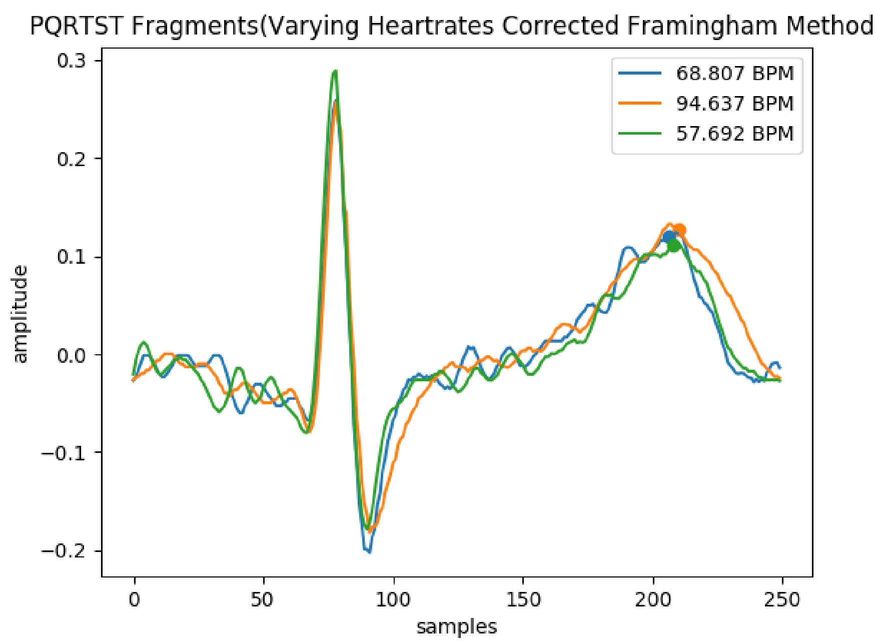


Figure 2.10: After applying Framingham's formula and re-sampling, the corrected QT segment is shown to be normalized regardless of the heart rate.

representation of the signal. The next issue then follows: how many segments need to be attached together in order to form an accurate frequency domain representation of the signal?

In order to get an accurate representation of the ECG signals in the frequency domain, the signals need to be adjusted slightly. First, the signals need to be concatenated to produce a more accurate frequency domain representation. The simplest way to accomplish this is to attach the end of one segment to the beginning of another. This creates a time domain signal with a length of n beats. Next, the FFT of multiple signals needs to be performed. In order to accomplish this a rectangular window is formed around n beats, where n is a hyper parameter that represents how many beats are in this rectangular window. Once the FFT is taken the signal is shifted by one beat until the window has gone through the entire concatenated ECG signal.

As seen in Figure 2.11, the FFT of a rectangular window is successfully taken. Looking closely at the figure, spikes can be seen at a consistent frequency through out the feature space. These spikes are caused by the discontinuity in the signal when the time domain ECG segments are concatenated. However, since all of the segments are of the same length these spikes will be consistent throughout all of the segments so during classification and feature space reduction these consistencies across all of the signals will be resolved. The size of the rectangular window is tuned during training and the classification accuracy is checked on both the training and validation set. This is described in greater detail in further chapters.

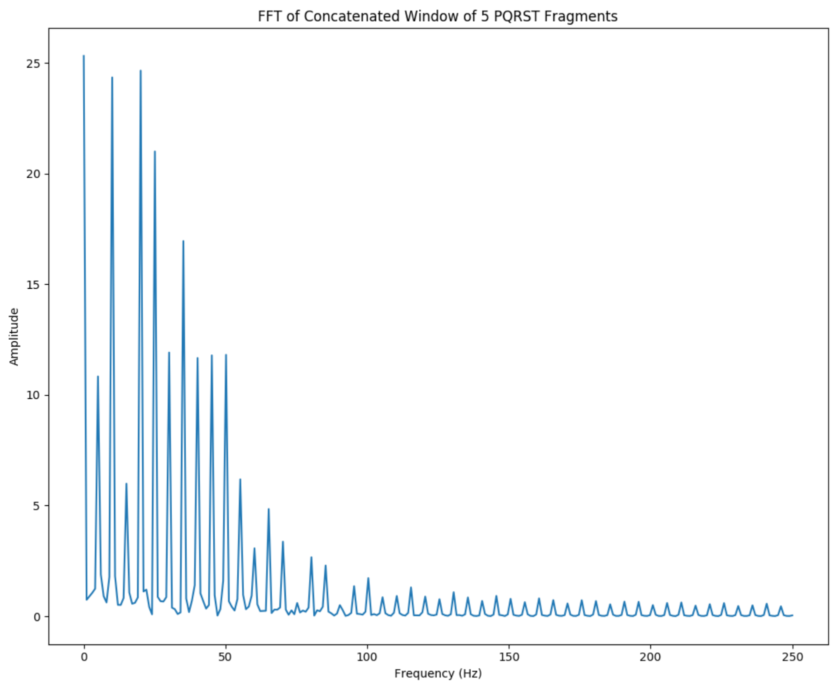


Figure 2.11: Concatenating the independent PQRST fragments creates aliased frequency spikes due to the discontinuity that is introduced during concatenation.

Chapter 3

Feature Space Formation

3.1 Background

3.1.1 Training, Testing, and Validation Sets

This chapter introduces the methods by which this research organizes the feature space into various sets or groups of data that are used for classification. These sets are the training, test, and validation sets. The purposes of these sets are as follows:

- Training Set:

  Used to train the classification algorithms. This set is formed from a certain percentage of the records for each individual. It is important that this set is separate from the test set so the produced results of the classification algorithms are not classifying data that the algorithm has been trained on. In other words, the final results will be tested on records that the algorithm has never seen before.

- Validation Set:

Since it is not good practice to develop the algorithm to fit the test set this research uses a validation set. The validation set is similar to the test set except it is used to test the hyper parameters and trained weights before testing on the test set. During training, especially while training neural networks, multiple iterations through the entire training set data will occur. These iterations are known as training epochs. In between each epoch, the training accuracy is determined by testing the current weights or status of the classification algorithm on the validation set. This information can then be used to determine whether the algorithm should continue to be trained and tweaked, or if it is ready for evaluation. In this research, the validation set is formed by removing a small percentage of the training set so that the validation set contains at least one record from each of the individuals.

- Test Set:

  The test set is used to perform a final evaluation of the classification algorithm. As mentioned earlier it is important that the test set remains separate from the training and validation sets, such that the algorithms are evaluated on their ability to classify data that they have not yet seen. The test set is a smaller percentage of all of the records than the training set and spans all of the individuals so that there is more than one record from each individual contained in the set. The test set comprises of approximately 30% of all of the total records, while the training set is 70% of the total records.

3.1.2 Principal Component Analysis

In this research, classification algorithms were evaluated on both reduced and unreduced feature spaces in order to determine what the best approach to classifying the ECG data is. This means that some form of feature space reduction needs to be performed. This research employs a technique called Principal Component Analysis or PCA to reduce the feature space from a dimension D to a dimension M where $M < D$. With PCA, it is possible to determine the importance of certain features within the D dimensional feature space and then use this information to transform or project the feature space into an M dimensional one.

Principal Component Analysis can be thought of as an orthogonal projection of a higher dimension feature space onto a lower dimensional linear space or principal space [4]. This is done in such a way that it minimizes a cost function that is the mean squared error between the feature space and their projections [4]. To explain this, consider a data set X of t n-dimensional feature vectors so that X is a $t \times n$ matrix. The goal is to develop a map from X to a new feature space that we will label Φ that is of a lower dimension D. This means that Φ is a $t \times d$ matrix. In other words, Φ will be a set of t d-dimensional feature vectors where $d < n$. We also define a matrix $V_{d\times n}$ that is a map that takes the matrix Φ back to the reconstructed n-dimension feature space $\hat{X}$. To perform PCA, in this research we considered the optimization problem of minimizing the reconstruction error of a rank d approximation of the original data matrix. This can be written as follows:

$$min_{\Phi,V} \sum_{i=1}^{t} ||X_i - \hat{X}_i||_2^2$$

where

$$\hat{X} = \Phi_{txd} V_{dxn}$$

Minimizing the above cost function finds the optimal rank d approximation of our data matrix X [12].

ECG Data Set Record Allocation

	domain	People/Classes	PQRST Segments	Total Records	Avg Length Per Record (seconds)
Training Set	Frequency		3146		
	Time	84	3723	206	20 – 30
Test Set	Frequency		858		
	Time	84	1161	104	20 – 30
			Total	310	

Table 3.1: Records and PQRST segments used for the training and test data sets. The unaltered ECG records were provided by PhysioBank's ECG database [7].

3.2 Data Sets

This section discusses how we went from ECG records, to individual PQRST segments, and then to three distinct and separate data sets that are labeled the training, test, and validation sets. This section details how many records are designated for each of the data sets and how strict we were at removing outliers. Although the same algorithm can be used for all the data sets, improved accuracy was seen when we imposed stricter outliers conditions to the validation set, and less strict conditions on the training set.

To start, the records for all individuals are split into individual records and sent to the various data sets. These totals are seen in Table 3.1.

3.2.1 Training Set

When forming a training set it is important to maximize the amount of data that is available. After splitting up the records between the training, test, and validation set, the PQRST segments need to be extracted and formed into a matrix defined in the final formation section given by Equation 3.1. This is done by following the steps outlined in chapter 2 : Data preprocessing of this thesis. The step in the data preprocessing where outliers are removed from the data set removes the most segments. Although this is an important step in removing "bad" ECG recordings, it is also a value that can be tuned.

As explained before, an outlier was determined by first calculating a mean segment from all of the PQRSTU segments within the record, and then calculating the euclidean distance of each record from that mean segment. We can then remove the segments by selecting a certain threshold in which any segments with a distance greater than this threshold are subject to removal. If this threshold is too strict, then we will remove too much of the data and see a decreased accuracy for two reasons. The first being that there is not enough data to train. The second is that there is not enough diversity in the data. Although training accuracy will see benefits due to the lack of diversity; when it comes to testing the classification algorithm on the testing set the algorithm will have been over trained and will not be as accurate. A higher threshold is chosen for the training set allowing more noise to exist in the data but at the same time achieving a higher training set diversity.

3.2.2 Validation Set

For the validation set, the same process used in forming the training set was performed including the same outlier removal threshold. In order to understand why this was done, it is important to think of the purpose of the validation set. The validation set is formed from a subsection of the training set. When training the data, we used the validation set to figure out whether the algorithm needs to be trained more or less. This is better than just testing the training accuracy on the training set itself because it does not show that the algorithm is good at classifying diverse data. For these reasons, the validation set employs the same outlier removal threshold as the training set formation.

3.2.3 Test Set

Lastly, the test set must be formed. The test set is meant to be completely separate from the training set and will be used to give the final results of the classification algorithms. Although it is important to maintain good diversity in the test set, significantly less data is needed in the test set than the training set. This means that when forming the test set, stricter thresholds can be used to remove most outliers. Although this is going to remove some "good" data there will be at least several PQRSTU segments from multiple records of each individual, so classification accuracy is fairly tested.

Unlike the training and validation sets, the test set's outliers are not removed via threshold. Similar to the training and validation sets, the euclidean distance from the mean PQRSTU segment in each record was calculated. However, instead of removing based off a threshold, the six smallest distances from the mean are chosen for each record. This is not done for the training set since it would remove too much data to accurately train without over-fitting the models. Six records were chosen to stay consistent with Lugovaya's work so that a good comparison can be made [13].

3.3 Final Formation

After all of the data sets have been processed, they are formed into a matrix that can be used in the actual algorithms. First, these records need to placed into a form that will be used to train the algorithms. These are $M \times N$ matrices in which M is the dimension of a record and N is the number of records in the set. After performing the data preprocessing techniques, the records have dimensionality of 250, a set will be 250xN dimension (unreduced). This can be seen as

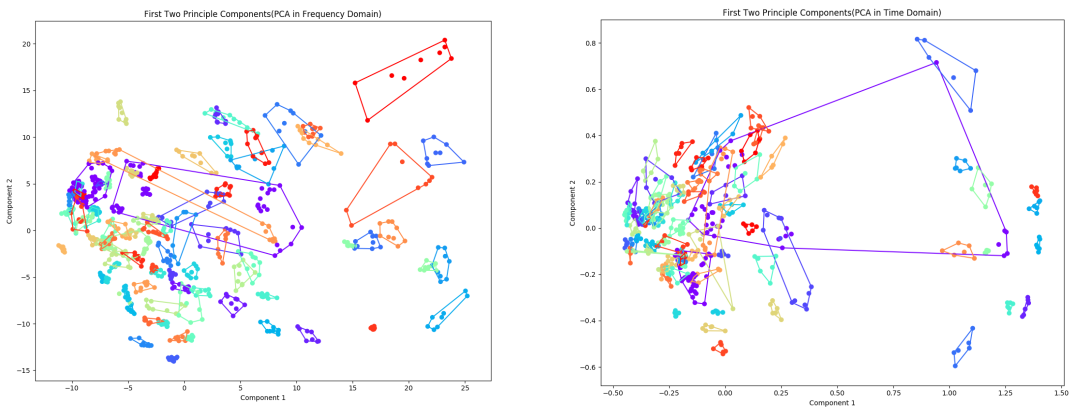
$$X = \begin{bmatrix} x_{0,rec_1} & \cdots & x_{0,rec_m} \\ x_{1,rec_1} & \ddots & x_{1,rec_m} \\ \vdots & \ddots & \cdots \\ x_{m-1,rec_1} & \ddots & x_{m-1,rec_m} \\ x_{m,rec_1} & \cdots & x_{m,rec_m} \end{bmatrix} \tag{3.1}$$

3.3.1 Unreduced Feature Space

Since this research classifies data in both the frequency and time domain, it is important to formulate the data matrix defined in Equation 3.1 in both the time and frequency domain. To simplify the formulation, the data matrices were chosen to have the same dimension. Both the frequency and time domain feature spaces form a $250 \times N$ data matrix.

3.3.2 Reduction Through Principle Component Analysis

This research tries reducing through PCA to various dimensions to see how it will effect the results of the classification discussed further in the chapter 5: Results/Discussion of this thesis. For visualization purposes, it can be seen that performing PCA helps show that classification of individuals is possible. This can be seen in Figure 3.1a and 3.1b, where two of the principal components are plotted against each other. Each color represents a separate individual. For visualization a box has been drawn that encloses all of the points within a certain class. As seen in the figure, groups become separated from each other. However, there are still some groups that are remaining closer together and share the same component space as other individuals. This is only in two dimensions so as more principal components are added to the feature space, these groups further differentiate from each other. The topic of added dimensions increasing accuracy is covered in more detail in the results chapter when reductions of various sizes are used to find an optimal solution. As seen in the figures, the frequency domain components appear to be more evenly spaced from each other. This is of course just a visualization technique but it shows that the frequency domain presents a separate set of data that can be utilized.



(a) Two Principal Components of the frequency domain data set with boxes drawn around the individual classes.

(b) Two Principal Components of the time domain data set with boxes drawn around the individual classes.

Figure 3.1: Seen in these figures is a separation between individuals with two principal components. The frequency and time domains produce similar results but different sets of information. This suggests that with more dimensions the classes will separate from one another

Chapter 4

Classification

4.1 Background

4.1.1 Linear Discriminant Analysis

In this research, Linear Discriminant Analysis is used as one of the classification techniques to identify individuals. In order to understand what linear discriminant analysis is, one must first understand what a discriminant function is. To do this, consider an input vector x that is part of one of K classes. A discriminant function is a function that takes this input vector x and assigns it to one of these K classes. For linear discriminant analysis, this means that the discriminant function is linear. Mathematically, this means that it is taking a input vector of a dimension D and projecting into a different dimension N. A simple linear discriminant function can be seen in Equation 4.1. For example, when there are two classes C_1 and C_2, this linear discriminant function will project an input vector of dimension n to a single dimensional point. Then a y_0 is chosen such that, if $y(x) \geq y_0$, then the input vector x is part of class C_1. Otherwise, it is classified as class C_2. When increasing to a K-class discriminant function, K linear discriminant functions are created and of the form $y_k(x) = w_k^T x + w_{k0}$. A input vector X is then classified if $y_k(x) > y_j(x)$ for all $j \neq k$. Meaning that the decision boundary for a K class discriminant function is K-1. So if there are two classes this means that the decision boundary is a line. If there are three classes, then a plane. More generally, it is a K-1 dimensional hyper plane [4]. Linear Discriminant Analysis refers to the act of calculating the weight vector w such that the linear discriminant function maximizes class separation while minimizing within class covariance to yield the output

$$y(x) = w^T x + w_0 \tag{4.1}$$

where w is a weight vector and w_0 is a bias.

There are multiple methods that can be used to solve for the weights of this Linear Discriminant function. However, in this research the Fisher Linear Discriminant method is used [4]. This section outlines how this method is performed and what it aims to accomplish.

To start, consider a classifier that seeks to classify an input vector x of dimension D into K classes. Also consider a set of linear discriminant functions without the biases as $y_k(x) = w_k^T x$. This creates the linear discriminant function that is a vector y, which can be seen in

$$y(x) = W^T x \tag{4.2}$$

where W is a matrix of weight column vectors w_k.

Fisher's linear discriminant analysis aims to accomplish two things:

- Maximize the separation between class means in order to maximize the separation between classes.

- Minimize the in class covariance in order to minimize the overlap between different classes.

To do this we need find both the within class covariance, defined as S_W, and the between class covariance, defined as S_B. The within class covariance matrix can be defined as follows:

$$S_W = \sum_{k=1}^{K} S_K \tag{4.3}$$

where

$$S_K = \sum_{n \epsilon C_k} (x_n - m_k)(x_n - m_k)^T \tag{4.4}$$

and

$$m_k = \frac{1}{N_k} \sum_{n \epsilon C_k} x_n \tag{4.5}$$

The between class covariance is defined as

$$S_B = \sum_{k=1}^{K} N_K (m_k - m)(m_k - m)^T \tag{4.6}$$

where the total mean of the data set is given by

$$m = \frac{1}{N} \sum_{n=1}^{K} N_k m_k \tag{4.7}$$

The cost function is then formed as a function of the within-class covariance and the between-class covariance. This can be accomplished by maximizing the ratio between the projected in class covariance and the between class covariance. This cost function can be seen in Equation 4.8. Once this function is maximized through various means, the weights are used in the linear discriminant function to classify the data.

$$J(W) = Tr\{(WS_WW^T)^{-1}(WS_BW^T)\}. \tag{4.8}$$

4.1.2 Feed-forward Neural Network

In this research, one of the classification techniques utilized is a feed-forward neural network. A feed-forward neural network consist of multiple layers: an input layer, n number of hidden layers, and an output layer. Each of these layers contain a certain number of nodes or perceptrons and weights. Each of these perceptrons contain an activation function that can be chosen by the user, the most common of which is the sigmoid function. However, in this research other activation functions are considered. The connections between each of the layers can be seen in Figure 4.1. In this figure, the connections between the layers have their own weights which will be trained while training the network. The input layer is where our feature vector X is fed into the network and the output layer denoted y is the output of the network. The output layer has the number of nodes that will be classified, and if the value is above a certain threshold then that node is the class that the network has decided.

A feed-forward network functions by cascading layers of perceptrons which contain activation functions. An activation function is a function that takes in an activation and outputs a value. Consider the first layer of the feed-forward neural network. To start, there is an input vector X of dimension D $(x_1, x_2, \cdots, x_D)$. Now consider that the first hidden layer after the input layer has M nodes. In this case the activation for each node in the hidden layer is given by Equation 4.9.

$$a_j = \sum_{i=0}^{D} w_{ji}x_i + w_{j0} \tag{4.9}$$

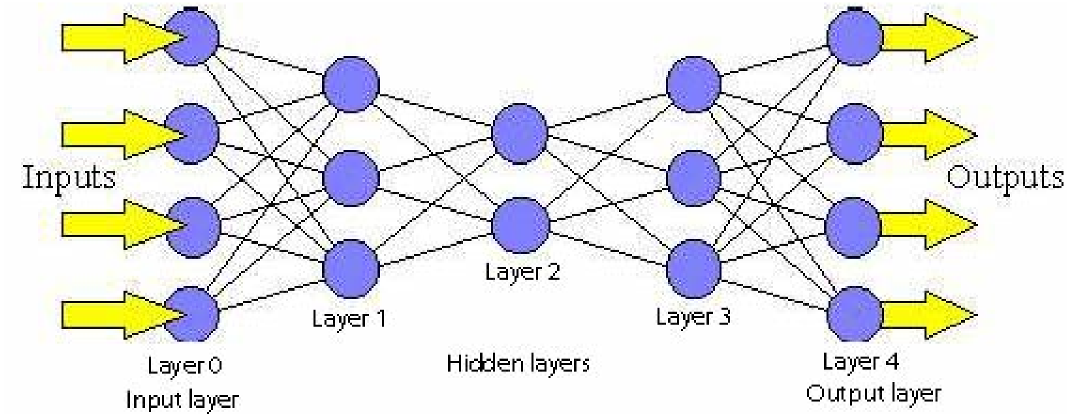


Figure 4.1: Diagram of a basic feed-forward neural network architecture with 3 hidden layers. Courtesy of https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Architecture/feedforward.html

where $j = 1, 2, \cdots, M$, w are the weights to be trained , and w_{j0} is the bias.

There is now an activation for each perceptron in the hidden layer; these activations can be put through an activation function to determine their output:

$$h(a_j) = z_j \text{ or } h(a_j) = y_k \tag{4.10}$$

when this is the output layer.

The activation functions are non-linear differentiable functions that can take different forms. The activation functions need to be differentiable because during training they are differentiated when trying to descend to a minimal error solution when back propagating the error. Two activation functions that are commonly used, and are used in this research, are the sigmoid and tanh activations functions. These can be seen in

- tanh:

$$h(\alpha) = \frac{e^{\alpha} - e^{-\alpha}}{e^{\alpha} + e^{-\alpha}} \tag{4.11}$$

- sigmoid:

$$h(\alpha) = \frac{1}{1 + e^{-\alpha}} \tag{4.12}$$

Note that these are mathematical equations mapping the outputs of one layer to the inputs of the next; the equations can be recursively called until the output layer is reached. These equations are how the feed-forward neural network moves from an input layer to an output layer.

In order to train this neural network, a procedure known as back propagation is used to find the gradient of the error with respect to the weights of the network. Then some form of gradient descent is used in order converge upon a minimal error solution. The steps of back propagation are included below [5]:

- Feed a vector X into the input layer and through the corresponding layers calculating the activations and output of all of the layers.

- Find the error between the target value and the calculated output for each output node.

- Back propagate this error, finding the error for each corresponding node within the hidden layers all the way to the input layer.

- Evaluate the partial derivatives of the total error of the network with respect to each layer's weights using these back propagated errors.

- Update the weights using gradient decent or another optimized minimization method.

4.2 Classification Using Linear Discriminant Analysis

Linear Discriminant Analysis does not have as many hyper parameters to tune as a feed-forward neural network. When training the LDA algorithm a parameter that was looked into is the dimension of the reduced feature-space. In the previous section, using PCA to reduce the dimensionality of our feature-space was discussed. In Lugavaya's research [13], a dimension of 30 was used. Our research asks the question of "why 30?" For this research's LDA classifier multiple different dimensionality reductions were used and tested. These are discussed thoroughly in the results chapter of this thesis.

4.3 Classification Using a Feed-forward Neural Network

When training a neural network, many parameters need to be tuned in order to create an optimal solution. These parameters are known as hyper parameters. The hyper-parameters are determined by initially using intuition, then watching how the network reacts during training and choosing or tuning a parameter to change based on what has been observed. The hyper-parameters used for tuning in this research are as follows:

- **Learn Rate**:

  Between each batch there is a form of gradient decent to update the weights of each of the layers. The learn rate determines how large of a distance along the gradient the weights will be changed.

- **Batch Size**:

Instead of training on the entire data set at once or on one PQRST segment at a time, batches can be formed. These batches are chunks of randomly grouped PQRST segments from the entire data set.

- **Training Epochs**:

  An epoch refers to the amount of data that the network has been trained on. One epoch means that the Neural Network has seen all of the training data once. In the case of batch gradient decent training, one epoch means it has run through all of the randomly formed batches, then new randomly formed batches are created so that the next epoch of training can occur.

- **Percent Dropout**:

  Dropout is a technique used to attempt to prevent overfitting, which is a condition that occurs when a classifier, in this case a feed-forward neural network, has converged upon a solution that is too closely related to the training set and not a generalized model. When the classifier is used on the test set, good results will not be achieved. Dropout attempts to mitigate overfitting by randomly choosing a percentage of the nodes to remove from a layer of the neural network.

- **Number of Hidden Layers**:

  As seen in the feed-forward neural network diagram, Figure 4.1, between the input and output layers there are several layers known as the hidden layers. One way that a network can be tuned is by increasing or decreasing the number of hidden layers.

- **Size of each Hidden Layer**

  Each of the hidden layers mentioned above can be adjusted by either giving them more or fewer nodes.

This research is classifying a group of 84 people with an input vector of length 250. The number of output nodes is 84, a node for each individual. The number of input nodes is 250 because the input vector has 250 features. The number of hidden layers and nodes in each of the hidden layers is discussed in the results section of this thesis as well as the values of the other hyper-parameters that are tuned.

4.4 Majority Vote Classification

Lastly, this research uses a majority vote classifier to finalize the classification of the heartbeat. This is a similar technique that is employed in Lugovaya's research and is included in this research for comparison. When classifying an individual, a 20 to 30 second ECG record is taken from the test set. After using the data processing methods for the test set that are outlined in the Feature Space Formation chapter 3 of this thesis, there are six PQRST fragments. Each of these fragments are then classified using one of the mentioned classification techniques in either the frequency domain, time domain, reduced, or unreduced feature space. Now each of the PQRST fragments are assigned a class. Initially the class could be determined by the most classifications, however this would not be ideal because if an individual is not part of the set of classes then there will be a false positive classification. Instead, the majority vote classifier assigns the entire record to a class if and only if three or more of these beats belong to that class. This means that the records can either be classified to a class or can be classified as unknown.

Chapter 5

Results/Discussion

This chapter focuses on the results of all of the different classification and reduction techniques used in this research. In addition, it discusses how and why the hyper parameters were tuned. Each classifier's results are shown in the frequency and time domain with their equivalent algorithms in a reduced feature space kernel. After each individual classifier is discussed, this chapter compares each of the algorithms and discusses the significance of the findings.

5.1 Linear Discriminant Analysis

After the training data has been processed, there are multiple PQRST fragments that the linear discriminant analysis model needs to be fitted to or trained with. In order to train the linear discriminant analysis classifier, the steps outlined in the previous chapter were performed. This involves minimizing the cost function that maximizes class separation while minimizing in class covariance using the training set. After this step has been completed, the model has been trained. This means that there now exists a set of weights that transforms our data into a lower dimension subspace that can be used for classification.

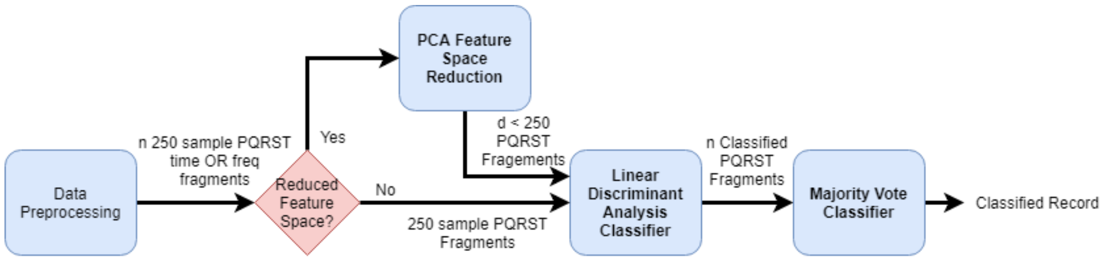


Figure 5.1: Flow chart of the classification of individuals ECG records using the Linear Discriminant Analysis Classifier.

Finally, individual's records are classified by first splitting up the test set records into the individual PQRST segments and using the weights calculated to classify each individual PQRST segment within the record. After this has been completed a record's class is determined by assigning a class based on the class that occurs in a majority of the PQRST segments. For example, if there are six records and three of these records belong to individual A and two records to individual B and one to individual C, then the record would be classified as belonging to individual A. This classification is done for all of the records in the test set. For the classification of the reduced feature space, the process is the same except the PQRST segments are reduced using the PCA before being classified. A flow chart of this process can be seen in Figure 5.1

5.1.1 Time Domain

The first classifier data that is investigated is the time series linear discriminant analysis classifier. This classifier was tested using multiple feature space sizes from one dimension to no feature space reduction at all. The percent accuracy was calculated on the test set and plotted against the feature space dimension. This can be seen in Figure 5.2.

As seen in Figure 5.2, using a reduced number of dimensions yielded low ac-

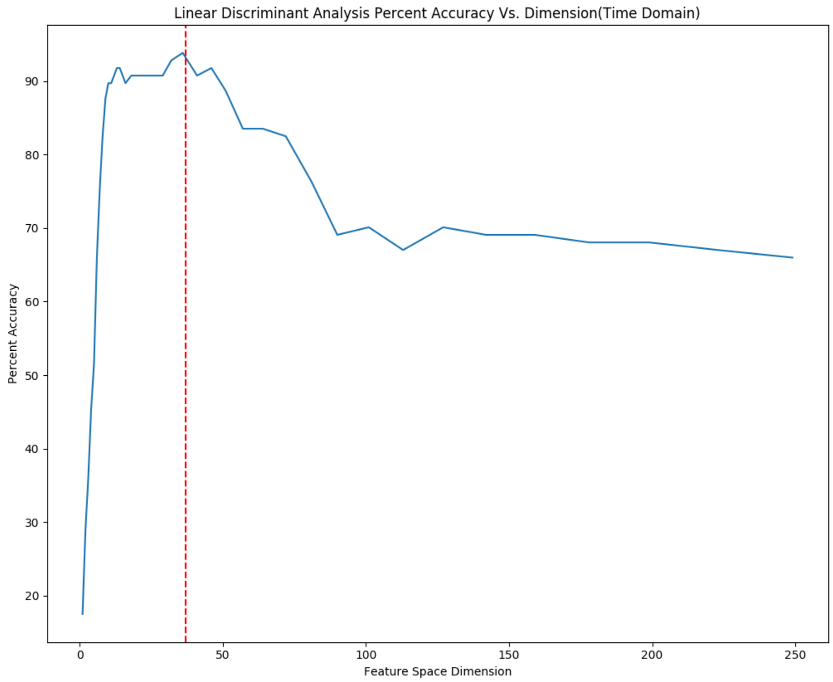


Figure 5.2: As the feature space dimension increases, there is an increase in accuracy until around $D = 50$, at which point the accuracy declines. This means that there is an optimal dimension that yields the highest accuracy results at $D = 40$ and can be seen as the vertical dashed red line.

curacy results. As the dimension number in the feature space increases, better results are achieved until the dimension gets too high and diminishing accuracies are observed. With the feature space reduced to 40 principal components the highest accuracy is achieved at 93%. This dimension reduction and percent accuracy is around the same values achieved in Lugavaya's research, verifying Lugavaya's the findings.

5.1.2 Frequency Domain

After classifying the time series data, the frequency domain feature-space was investigated, where a similar procedure was performed. In Figure 5.3, the classification accuracy can be seen with respect to the feature space dimension used.

As seen in the time series classification, a peak accuracy of the records is in between the unreduced and single dimensional space. Unlike the time series the frequency domain classification occurred at a higher dimension, indicating that the frequency data requires more dimensions to separate classes. The peak accuracy of this classifier was 82%, which is a large drop in accuracy compared to the time series data.

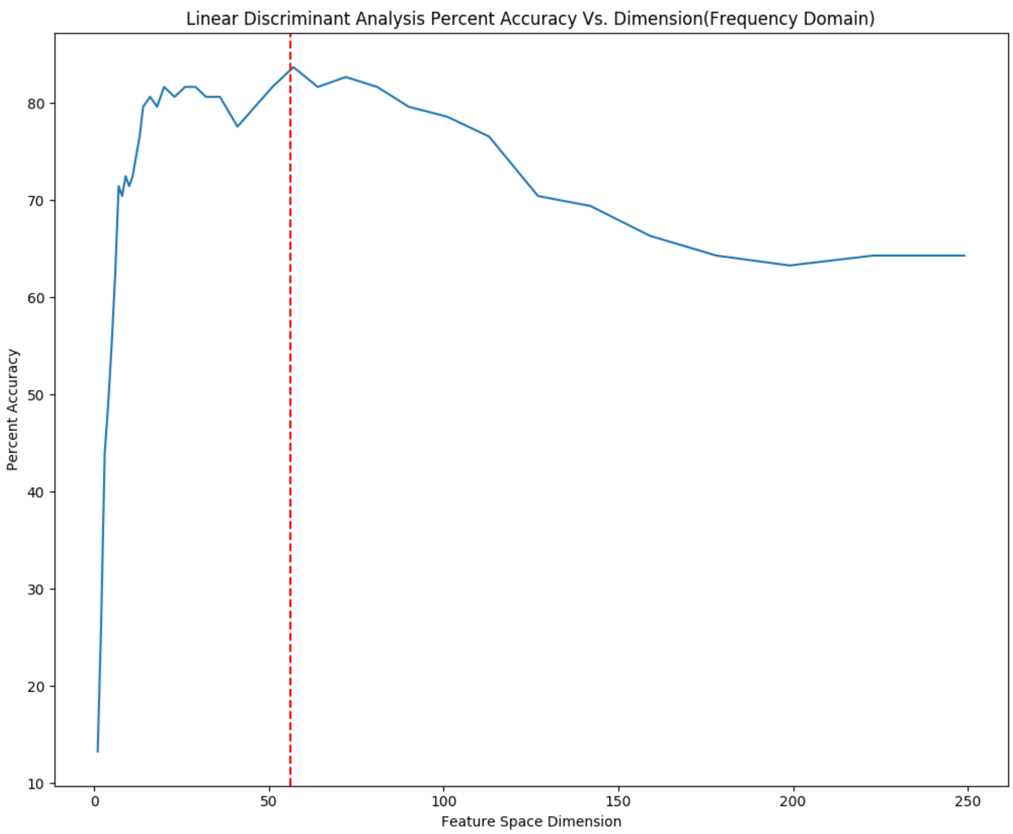


Figure 5.3: Similar to the time domain classification the LDA classifier has an optimal dimension where the accuracy is maximized. This occurs at dimension $D = 56$ and can be seen as the vertical dashed red line.

5.2 Feed-forward Neural Network

Similar to the Linear Discriminant Analysis classifier used in the previous section, the feed-forward neural network is used to classify individual PQRST fragments of the entire ECG record. These fragments are sent to the majority vote classifier so that the entire record is assigned to a single class. This process can be seen in Figure 5.4. This section covers the results of the training and testing of the feed-forward neural network and majority vote classifier in both unreduced and reduced time and frequency domain feature spaces.

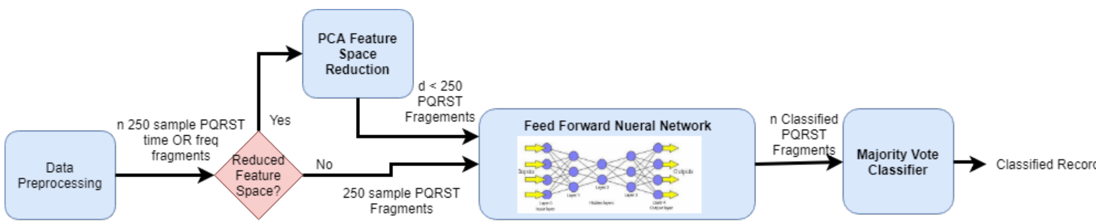


Figure 5.4: Flow chart of the classification of individuals ECG records using the feed-forward Neural Network and Majority Vote Classifier.

Before showing the results of the classifier it is important to discuss the loss function that is used. For all of the feed-forward neural network the loss function that was used is ***log loss*** or ***cross-entropy*** loss. This is defined in Equation 5.1

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log p_{ij} \tag{5.1}$$

where N is the number of samples and M is the total number of possible classes. p_{ij} is the probability of assigning a certain class and y_{ij} is a one or zero indicating whether the classification is correct.

5.2.1 Time Domain

After many iterations, the parameters of the feed-forward neural network that yield successful results and a converged solution used the hyper-parameters seen in Table 5.1. Using these hyper-parameters created the results during training seen in Figures 5.5 and 5.6. In Figure 5.6, the loss is evaluated during each training epoch on the training and validation set and can be seen dramatically decreasing and converging upon a solution with little ***log loss***. One hundred total epochs are shown in this Figure; however, in an attempt to not overfit the data the training process is stopped early based on the values seen. The accuracy can be seen in Figure 5.5, where it eventually converges upon a roughly constant accuracy for both the training and validation set. This convergence occurs around 80 epochs, meaning 80 training epochs were chosen for the unreduced time series classifier. When training with these values, the training set accuracy and loss observed can be seen in Table 5.2.

Time Domain FFNN	
Hyper-parameter	**Value**
Learn Rate	0.01
Batch Size	10
Training Epochs	80
Percent Dropout	0.5
Hidden layers	1
Hidden layer 1 size	200

Table 5.1: Hyper parameters chosen for training the feed-forward neural network of unreduced time domain input vectors.

Training Set	
Training Accuracy	0.87
Logistical Loss	6.823

Table 5.2: Training set accuracy and loss at the end of training.

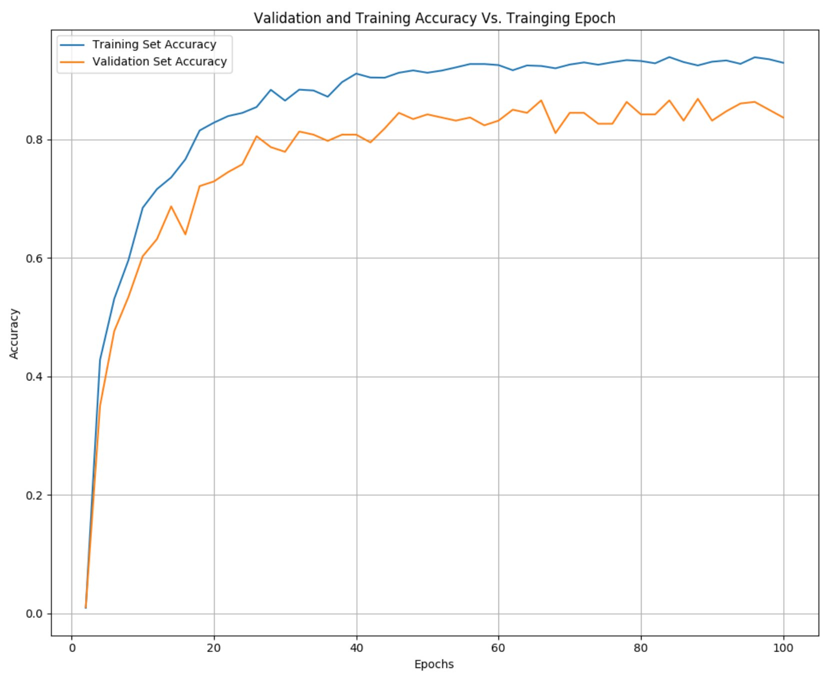


Figure 5.5: After several epochs the training and validation set converge upon a solution. The validation accuracy is lower than the training accuracy. This is because the neural network has not yet seen the validation set.

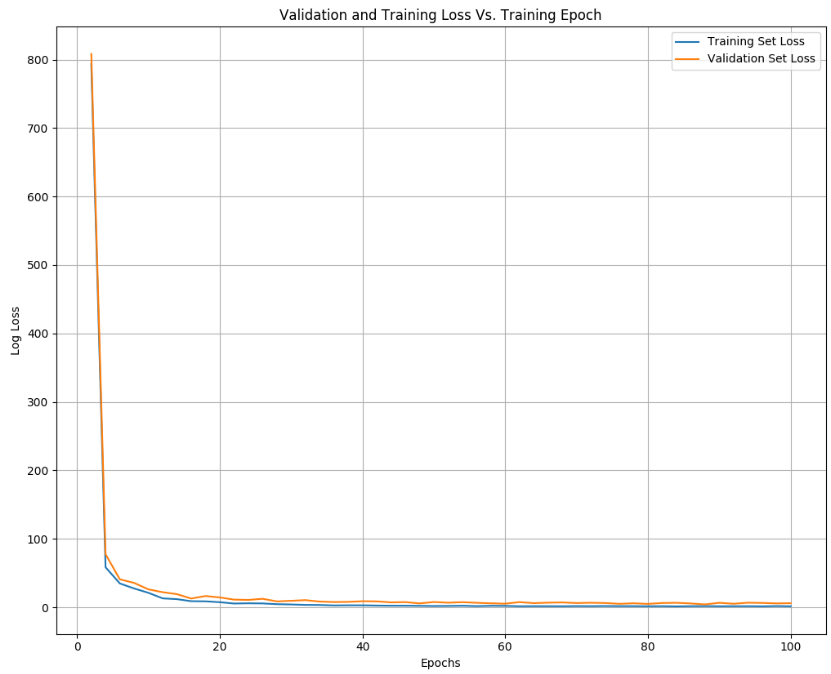


Figure 5.6: The loss of both the training and validation set drop quickly. The ***log loss*** converges in less epochs than the percent accuracy.

Lastly, after the training of the feed-forward neural network has been completed, the developed classifier was used with the test set to obtain results. This produces the final result to compare against the other classifiers used in this research. As mentioned earlier, the feed-forward neural network is used to classify individual PQRST fragments which are then sent to the majority vote classifier along with the other fragments from the same record to classify the entire record. Using this technique the results seen in Table 5.3 are achieved.

Test Set (clasifiying records)		Accuracy
Total Prediction	98	-
Correct Predictions	91	0.92857
Incorrect Predictions	7	0.71486

Table 5.3: Test set classification of entire ECG record using feed-forward neural network to classify PQRST fragments and a majority vote classifier to classify the entire record yields a 92% accuracy which is comparable to the the LDA classifier.

Next, the classifier was designed using the reduced dimension version of the time series feature space. While training the network, the dimension that yielded the best results was a dimension of $D = 50$. The hyper-parameters accompanying this can be seen in Table 5.4. The same general trend in training accuracies and losses using the unreduced feature space is seen when training with the reduced feature space. In Figure 5.7 and 5.8, the training and validation set losses and accuracies can be seen respectively.

Time Domain FFNN	
Hyper-parameter	**Value**
Learn Rate	0.01
Batch Size	10
Training Epochs	100
Percent Dropout	0.8
Hidden layers	2
Hidden layer 1 size	100
Hidden layer 2 size	100

Table 5.4: Hyper parameters chosen for training the feed-forward neural network of the reduced time domain input vectors.

After the training is complete, the test set data was evaluated using the trained classifier as well as the majority vote classifier to obtain the results that were compared against the other classifiers. For the reduced feature space time domain records, the results can be seen in Table 5.5

Test Set (clasifiying records)		**Accuracy**
Total Prediction	97	-
Correct Predictions	89	0.91752
Incorrect Predictions	8	0.82474

Table 5.5: Test set classification of entire ECG record using feed-forward neural network to classify PQRST fragments and a majority vote classifier to classify the entire record. The feature space is reduced to a dimension of $D = 50$ time series PQRST fragments. This shows that the test set classification accuracy is only slightly lower than its unreduced counterpart classifying at 91%.

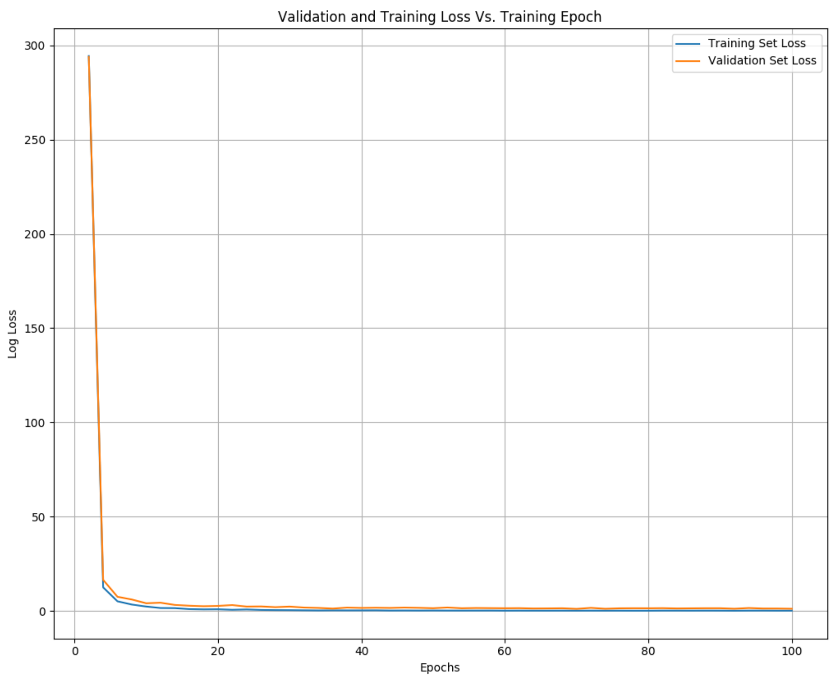


Figure 5.7: In the time domain, when dimension has been reduced to $D = 50$, the loss calculated at each epoch on the training and validation set can be seen to drop and converge to a solution.

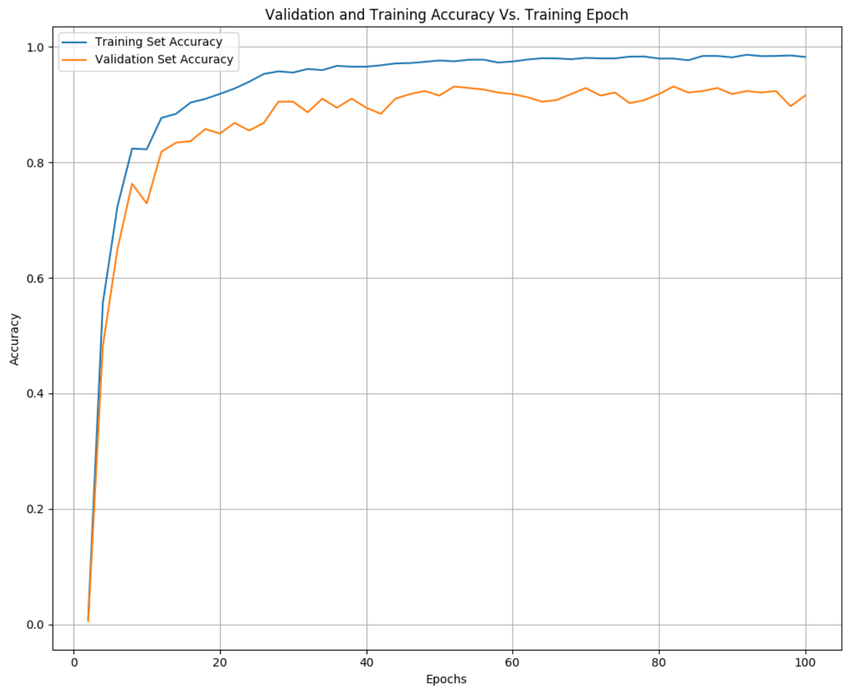


Figure 5.8: In the time domain, when dimension has been reduced to $D = 50$, the accuracy can be seen converging to a solution similar to in the unreduced feature space. This accuracy appears to be greater than that of the unreduced feature space, however the results on the test set and validation set show that this does not necessarily yield a better result.

5.2.2 Frequency Domain

After the time domain of the ECG records is inspected, the focus is shifted towards the frequency domain representation and its use within the feed-forward neural network classifier. This is the same process outlined in the time domain, except the hyper parameters have changed slightly. This is because it is a separate set of data, so the same hyper parameters do not necessarily yield the same best results. As with the time domain, the frequency domain classifier was tuned in both the unreduced and reduced forms of the feature space. To start, the unreduced feature space was inspected. After some tuning, the hyper parameters that yield the best results for the unreduced feature space can be seen in Table 5.6. The accuracy and loss of both the training and validation sets can be seen in Figures 5.9 and 5.10 respectively.

Time Domain FFNN	
Hyper-parameter	**Value**
Learn Rate	0.01
Batch Size	100
Training Epochs	500
Percent Dropout	0.5
Hidden layers	2
Hidden layer 1 size	200
Hidden layer 2 size	100

Table 5.6: Hyper parameters chosen for training the feed-forward neural network of the unreduced frequency domain input vectors.

Training Set	
Training Accuracy	0.987
Logistical Loss	3.4197

Table 5.7: Training set accuracy and loss at the end of training. The feature space is unreduced frequency domain windows of PQRST fragments.

Finally, after the classifier is trained, the test data was classified using both the

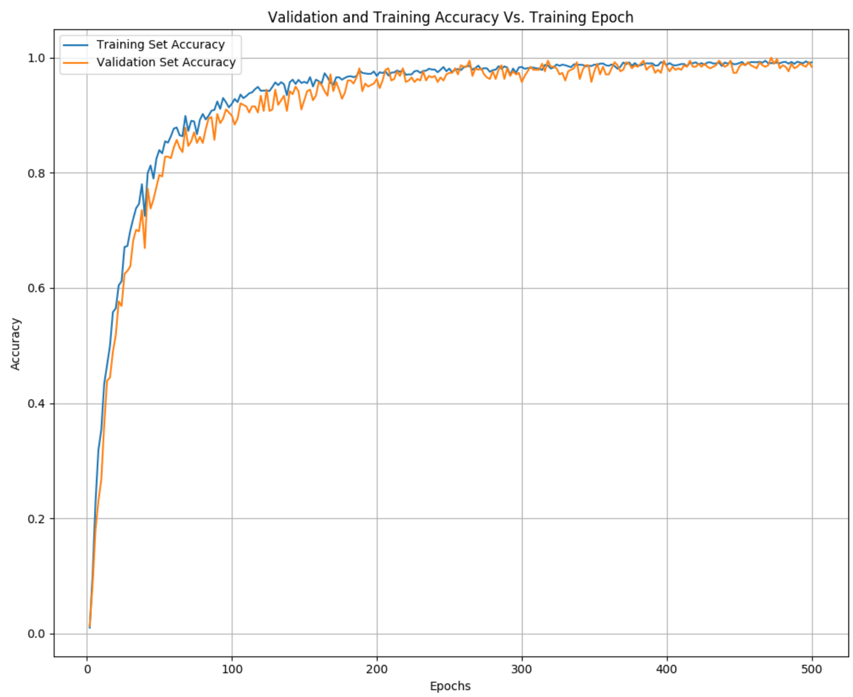


Figure 5.9: During training of the feed-forward neural network on the unreduced frequency domain feature space, it can bee seen that the validation and training set accuracy converge to a near 100% accuracy. However, as seen in the results this does not necessarily yield a higher accuracy on the test set. This means the data is most likely being overfit.

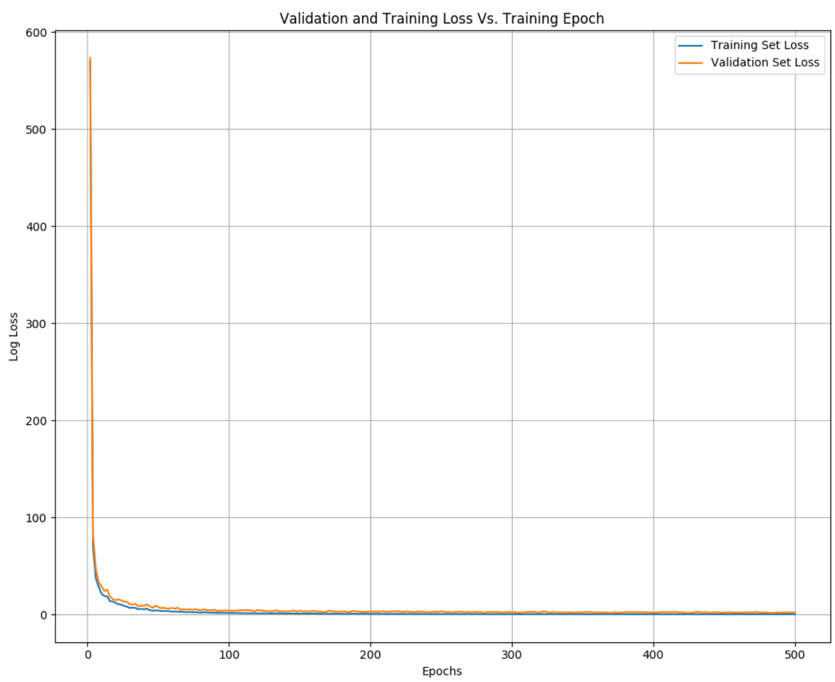


Figure 5.10: During training of the unreduced frequency domain feature space, the loss of both the training and validation set drop and converge quickly to a near 0 value. This shows that high accuracy can be obtained in the training set but due to overfitting the model does not perform as well on the test set.

feed-forward neural network and the majority vote classifier. Using the unreduced frequency domain representation of the ECG fragments, the classifier gave the results seen in Table 5.8. Although the training and validation set accuracy and loss performs better than in the reduced and unreduced time domain signals, the results on the test set are not proportionally better. Instead, the results are slightly worse indicating that the frequency data is more susceptible to overfitting.

Test Set (clasifiying records)		Accuracy
Total Prediction	97	-
Correct Predictions	85	0.87629
Incorrect Predictions	12	0.12371

Table 5.8: Test set classification of entire ECG record using feed-forward neural network to classify PQRST fragments and a majority vote classifier to classify the entire record. The feature space is unreduced frequency domain PQRST fragments. This shows that the unreduced frequency domain is classified with less accuracy despite a higher training accuracy.

After the unreduced frequency domain data was investigated, the reduced frequency domain data was used to create the feed-forward neural network classifier. For the frequency domain, after much training and tuning the dimension that provides the best results is a reduced dimension of $D = 70$, with the hyper parameters seen in Table 5.9. The training results using these hyper parameters can be seen in Figures 5.11 and 5.12. The training accuracy that was achieved can be seen in Table 5.10.

Time Domain FFNN	
Hyper-parameter	**Value**
Learn Rate	0.01
Batch Size	100
Training Epochs	200
Percent Dropout	0.5
Hidden layers	2
Hidden layer 1 size	80
Hidden layer 2 size	90

Table 5.9: Hyper parameters chosen for training the feed-forward neural network of the reduced to dimension of $D = 70$ frequency domain input vectors.

Training Set	
Training Accuracy	0.971
Logistical Loss	2.8013

Table 5.10: Training set accuracy and loss at the end of training. The feature space is reduced to dimension of $D = 70$, frequency domain windows of PQRST fragments.

Lastly, since the training has been completed it is necessary to send the test data through the classifier and subsequent majority vote classifier. The results achieved for the reduced frequency domain representation of the feature vectors can be seen in Table 5.11. Similar to the unreduced frequency domain classifier, the training and validation data is classified more accurately than the time domain; however, the test results are not similarly better.

Test Set (clasifiying records)		**Accuracy**
Total Prediction	97	-
Correct Predictions	83	0.85567
Incorrect Predictions	25	0.14432

Table 5.11: Test set classification of entire ECG record using feed-forward neural network to classify PQRST fragments and a majority vote classifier to classify the entire record. The feature space is frequency domain windows of PQRST fragments reduced to a dimension of $D = 70$.

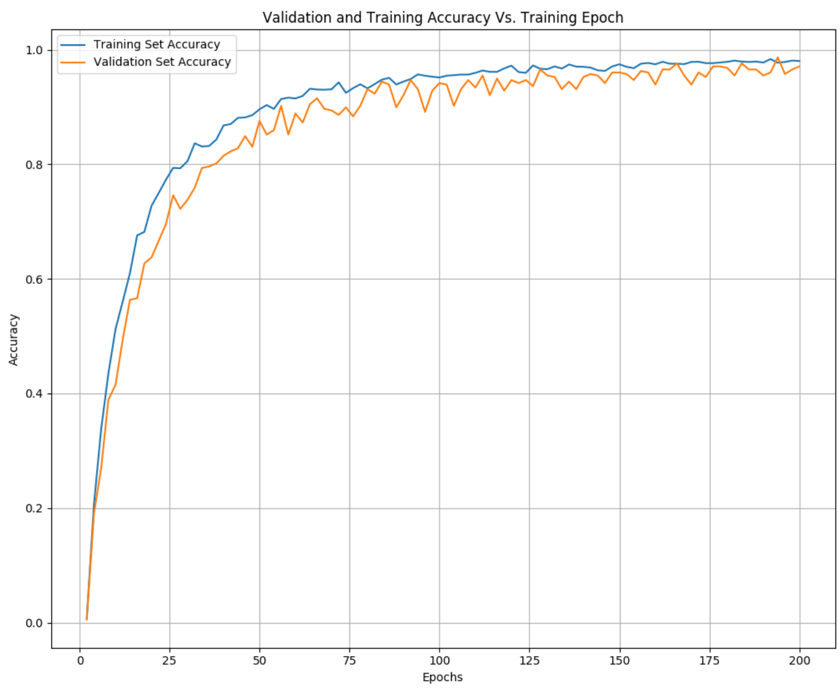


Figure 5.11: The accuracy of the reduced ($D = 70$) frequency domain feature space converges to a fixed accuracy. This accuracy is lower than its unreduced counterpart. This trend is also reflected in the overall accuracy on the test set.

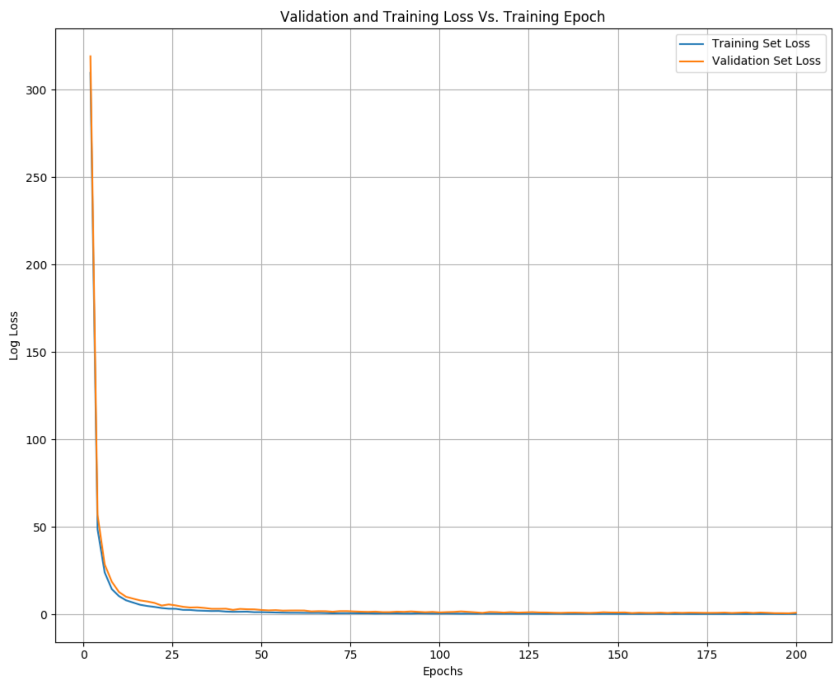


Figure 5.12: Similar to its unreduced counter part, the reduced to $D = 70$, frequency domain feature space shows a convergence in the loss with the loss being less than that of the time domain equivalent. However, when applied to the test set these results do not carry over.

5.3 Comparison

This section discusses and compares the different classifiers used by this research. It will also give possible reasons for the observed results. In addition to this, each type of classifier will be discussed individually. A comparison between all of the results is shown in Table 5.12.

Feed Forward Neural Network				
	Time Domain		Frequency Domain	
Results	D = 250	D = 50	D = 250	D = 70
Correct Predictions	91	89	85	83
Incorrect Predictions	7	8	12	14
Accuracy	0.92857	0.917525	0.876289	0.8567
Linear Discriminant Analysis				
	Time Domain		Frequency Domain	
Results	D = 250	D = 40	D = 250	D = 56
Accuracy	0.68	0.925	0.66	0.82

Table 5.12: Comparison of the accuracies of all the classifiers investigated. The best performing classifiers within the two types of classifiers is highlighted in green.

To start, look at the various scenarios of the Linear Discriminant Analysis classifier. In both the frequency domain and time domain, the unreduced feature space has equally unsuccessful results. Each has a peak classification occur around 50 to 70 principal components out of 250. This implies that the Linear Discriminant Analysis classifier has issues classifying based on a large amount of features. Feature space reduction is almost necessary in order to obtain successful results. This implies that the LDA classifier is over fitting to noise. As the feature space increases and moves closer to having no reduction, more unimportant or noisy features enter the system. There becomes a point that the added features do not provide any information that is unique to the individual. In other words, the reason why there is a peak in the data is that with little dimensions there is not enough unique information to separate the individual's ECGs. Then at a certain point the added features do not add anymore useful or unique information and therefore outweigh the more unique features, diminishing the accuracy. With the ECG data the time domain yields the best results with an accuracy of 0.92, which is comparable to the 0.96 obtained in Lugovaya's research using similar techniques [13]. The reduced frequency domain was not as successful as the time domain, with a peak accuracy of only 0.82.

In the case of the feed-forward neural network, the unreduced feature spaces in both the time and frequency domain produce higher accuracy classifiers than using their reduced feature space equivalents. The unreduced time domain feature space yields the best results with an accuracy of 0.928. The equivalent frequency domain classifier sees an accuracy of 0.87. This means that the unreduced time domain classifier gives results that are similar to the reduced Linear Discriminant Analysis classifier. In the case of the unreduced frequency domain feature space, the accuracy does much better than its equivalent LDA classifier but does not match that of the time domain.

5.4 Applications

In this section the practical uses or application of this research are discussed. This research has verified the work of Lugovaya in showing that ECG data can be used in the time domain to classify individuals using a LDA classifier, but has also shown that a feed-forward neural network can be used in both the time and frequency domain to achieve similar accuracy.

In order for this research's identification algorithm to be implemented in a real life situation there will need to a be a database of individuals ECG's. This database will need to be continually maintained. The ECG's that were used in this research were recorded with months of separation. This is important because it shows that the uniqueness of the individual's heartbeat remains unique over time. If another individual is added to the database the algorithms should be retrained appropriately.

One way that this data can be used is to enhance already existing security systems. It could be used as a way to further authenticate an individuals identity. Another possible application of this information can be that if an individuals regular ECG has been trained, it can be used to identify abnormalities in the individual. If the user's ECG cannot be identified when it previously could then this could be due to an abnormality in the ECG that could be a possible health risk. This topic would need further investigation.

Chapter 6

Conclusion

The feed forward neural network with an unreduced feature space produced similar results to the Linear Discriminant Analysis classifier with a reduced feature space in the time domain. However, in the frequency domain the feed forward neural network produced better in both the reduced and unreduced feature space than the LDA classifier. This is most likely due to the extra noise present in the frequency domain representation of the PQRSTU fragments. The extra noise introduced from the discontinuity during the concatenation of the corrected time domain signals is constant amongst all of the records. When using LDA to classify in the unreduced frequency domain feature space, the classifier over-fits to this noise because it outweighs the important information in the feature space. During feature space reduction, the noise is constant across all of the signals so it is effectively removed causing the resulting classifier to be more accurate. With the feed forward neural network we do not see the same effect. This means that with our data the feed forward neural network is more robust to the noise and is able to classify equally as well in both the unreduced and reduced feature spaces. This is because techniques like early stopping and dropout were used during the training of the neural network to prevent the classifier from over-fitting to the noise.

This research successfully showed that the frequency domain can be used in addition to the time domain to classify ECG records from a set of 84 individuals. Further research needs to be conducted to see whether or not this classifier can be expanded to a group that is larger than 84 individuals. The peak accuracy achieved in this research is 93%. With this result, ECG classification could be used to enhance already existing security devices to add another layer of protection.

Bibliography

[1] Neural networks. `https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Architecture/feedforward.html`. Accessed: June 2018.

[2] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[3] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, and Stanley HE. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals.

[4] Christopher Bishop M. *Pattern Recognition and Machine Learning*. Springer Science + Business Media, LLC, 2006.

[5] Christopher Bishop M. *Pattern Recognition and Machine Learning*. Springer Science + Business Media, LLC, 2006.

[6] Dubiel. Chen, Xie. Juslien. wfdb-python. `https://github.com/MIT-LCP/wfdb-python`, 2018.

[7] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). Circulation Electronic

Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[8] Bazett H. C. An analysis of the time-relations of electrocardiograms. 7:353–370, 1920.

[9] Moody. I, Silva. G. An open-source toolbox for analysing and processing physionet databases in matlab and octave. *Journal of Open Research Software*, 2(1):e27, 2014 (September 24).

[10] Cardiology Division St. Vincent Hospital Worcester MA. Survey of selected cardiologists for an operational definition of normal sinus heart rate. *American Journal of Cardiology*, 72:487–488, 1993.

[11] Alex Sagie MD, Martin G. Larson ScD, Robert J. Goldberg PhD, James R. Bengtson MD, and Daniel Levy MD. An improved method for adjusting the qt interval for heart rate (the framingham heart study). *American Journal of Cardiology*, 70:797–801, 1992.

[12] Farzaneh Mirzazadeh. Dimensionality reduction. *University of California, Santa Cruz*, January 2018.

[13] Lugovaya Tatians S. Biometric human identification based on ecg. Master's thesis, Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Sait-Petersburg, Russian Federation, June 2005.

[14] Medical Training and Simulation LLC. Ekg: Practical clinical skills. `https://www.practicalclinicalskills.com/ekg`. Accessed: 2017.