Evaluation of the Signal Quality of Wrist-Based Photoplethysmography

by

Nikhilesh Pradhan, BEng, Carleton University

A thesis submitted to the

Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Master of Applied Science in Biomedical Engineering

Ottawa-Carleton Institute for Biomedical Engineering

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, Canada

December 2017

The undersigned recommend to the
Faculty of Graduate Studies and Research
acceptance of the thesis

Evaluation of the Signal Quality of Wrist-Based Photoplethysmography

Submitted by Nikhilesh Pradhan, BEng, Carleton University
in partial fulfillment of the requirements for the degree of
Master of Applied Science in Biomedical Engineering

Thesis Supervisors
Professor Andy Adler
Professor Sreeraman Rajan

Chair, Department of Systems and Computer Engineering
Professor Yvan Labiche

Abstract

Detection of paroxysmal atrial fibrillation requires continuous cardiovascular monitoring due to its episodic nature. Such monitoring is impractical with electrocardiogram Holter monitors, which are the currently employed for ambulatory cardiovascular monitoring, but are cumbersome for prolonged use. This thesis studies monitoring using photoplethysmography (PPG) devices, which may be embedded into wristband devices which can be easily worn continuously. However, the quality of wrist-based PPG is highly variable, and is subject to artifacts from motion and other interferences. The goal of this thesis is to evaluate the signal quality obtained from wrist-based PPG when used in an ambulatory setting. Ambulatory data is collected over a 24-hour period for 10 elderly, and 16 non-elderly participants. Visual assessment is used as the gold standard for PPG signal quality, with Fleiss's Kappa being used to evaluate the agreement between raters. With this gold standard, 5 classifiers are evaluated using a modified 13-fold cross-validation approach. Based on this evaluation, a *Random Forest* quality classification algorithm is selected, with an accuracy of 74.5%. The algorithm is used to evaluate the ambulatory use of wrist-based PPG over a 24-hour period. Overall, it is found that data quality is high at night, and low during the day.

Acknowledgements

First and foremost, I would like to acknowledge all the support received from my family, without which none of this would have been possible.

I would also like to thank my supervisors, Professor Andy Adler and Professor Sreeraman Rajan, for their advice and guidance throughout the project. A special thanks to Marie Mulholland, at the University of Ottawa Heart Institute, for her assistance in obtaining ECG Holter data. I would also like to express my gratitude to Maryam Kaka and Jinny Lee for all their contributions to the data collection process. I would like to thank Kevin Dick for sharing his knowledge and passion for machine learning. I also would like to thank Tarek Nasser El Harake for his advice, which was invaluable for generating the data quality visualization (Figure 7.1).

Last but certainly not least, I would like to acknowledge my friends, Jaime Vives and Eshani Sharan for all their support over the past two years.

Contents

List of Abbreviations

κ	Fleiss's Kappa (Agreement Between Raters)
AF	Atrial Fibrillation
AUC	Area Under the Curve (For a ROC Curve)
CART	Classification and Regression Tree
DTW	Dynamic Time Warping
ECG	Electrocardiogram
GUI	Graphical User Interface
HR	Heart Rate
LED	Light-Emitting Diode
MLP	Multi-Layer Perceptron
P_i	Agreement Between Raters per Instance
PPG	Photoplethysmography
PSD	Power Spectral Density
qSQI	Fused Signal Quality Index

ROC Receiver Operating Characteristic

SNR Signal-to-Noise Ratio

SQI Signal Quality Index

SSA Singular Spectrum Analysis

SVM Support Vector Machine

Chapter 1

Introduction

1.1 Background

Long-term, continuous, non-invasive monitoring of cardiovascular activity in an ambulatory setting would enable early detection of cardiovascular problems, such as atrial fibrillation (AF), thereby allowing for medical interventions to preempt medical emergencies such as strokes and heart attacks. While undetected AF presents a high risk of complications, its pathogenesis is poorly understood before the condition presents itself clinically. Therefore, such monitoring would enable studies into the progression of AF, potentially leading to the development of predictive systems for cardiovascular risk assessment. Detection of paroxysmal AF is challenging as it is episodic in nature. The AF episode must occur while cardiovascular monitoring is underway to be detected. Thus, screening for paroxysmal AF episodes can only be reliably achieved through prolonged continuous cardiovascular monitoring, as short monitoring windows may not capture AF episodes.

However, long-term, ubiquitous monitoring is impractical with electrocardiogram (ECG) Holter monitors, the technology currently employed by medical institutions

for ambulatory use. ECG Holter monitors are considered cumbersome due to their requirement for electrodes and wires adhering to the thorax. In addition to being uncomfortable, it also places restrictions on the user, as the Holter system is sensitive to water damage; hence, users are unable to shower or swim while wearing the device. As a result, users are less likely to wear Holter monitors continuously for prolonged periods of time, especially if they are not currently suffering from a diagnosed heart condition. A potential alternative is photoplethysmography (PPG), which is an optical technology that is easily embedded into wearable devices such as wristbands, and is already included in many commercially available smartwatches. PPG detects local changes in blood volume by measuring the attenuation of light, which provides an indication of the cardiovascular activity. However, the technology is highly sensitive to noise-corruption, especially due to motion. Cardiovascular parameters derived from noise-corrupted data would be unreliable and potentially lead to inaccurate diagnoses. This necessitates the application of a classification system for the assessment of signal quality, to ensure that noise-corrupted signals are handled appropriately - either cleaned, or rejected prior to cardiovascular analysis.

This thesis evaluates the signal quality of wrist-based PPG used in an ambulatory, daily use setting. As signal quality of PPG is validated through visual analysis, an annotated dataset is produced, along with statistics analyzing the agreement between the raters who provided the annotation. A machine learning algorithm is developed for the classification of the quality of PPG signals, which is evaluated using the annotated dataset. The algorithm is then applied to 24-hour, ambulatory data collected from 26 participants to evaluate the quality of data obtained from wrist-based PPG. Thus, the thesis presents a holistic signal quality evaluation of wrist-based PPG technology for ambulatory use.

1.2 Objectives

Three primary objectives were set for this thesis:

- To evaluate wrist-based photoplethysmography for use in long-term ambulatory monitoring.
- To develop an algorithm for classifying the signal quality of wrist-based photoplethysmography.
- To analyze the reliability and consistency of visual assessment as a gold standard for photoplethysmography signal quality.

1.3 Contributions

Work conducted in this thesis resulted in the following major contributions:

1. Compared electrocardiogram and photoplethysmography signal quality for ambulatory use using visual assessment, with publication presented at the IEEE 2016 International Student Conference [1].
2. Developed algorithm for discriminating between noise-free and noise-corrupted photoplethysmography data, with publication presented at the IEEE IMS 2017 Medical Measurements and Application Conference [2].
3. Collected a detailed database of ambulatory wrist-based photoplethysmography data over a 24-hour period.
4. Conducted analysis of the rater agreement in visual assessment of the quality photoplethysmography signals.

5. Developed an algorithm for classification of photoplethysmography signal quality on a 5-class scale.

6. Evaluated the signal quality of photoplethysmography in ambulatory use over a period of 24 hours using the classification algorithm.

1.4 Organization of Thesis

An introduction into the problem addressed, the objectives and the contributions of this thesis are provided in Chapter 1. Chapter 2 provides a background into photoplethysmography technology, as well as an overview of quality analysis algorithms proposed in literature. Chapter 3 provides the data collection methodology employed to collect the wrist-based PPG data. Chapter 4 is an analysis of the variability and reliability of the gold standard obtained by visual assessment of the PPG signal by raters. Chapter 5 is an overview of the feature extraction, with details on the implementation and performance of each of the 71 features. Chapter 6 develops the classification algorithm, and discusses the feature and classifier selection approaches used to obtain the final algorithm. Chapter 7 evaluates the signal quality over a 24 hour period through the application of the algorithm developed in Chapter 6. Chapter 8 concludes the thesis, with an overview of the results obtained from the analysis.

Chapter 2

Background

2.1 Photoplethysmography

Photoplethysmography (PPG) is an optical technique for monitoring cardiovascular activity through the measurement of local changes in blood volume. In its simplest form, PPG consists of two components; a light source and a photodetector. The light source transmits light through the microvascular bed of tissue. A portion of the transmitted light is absorbed by the various constituents of the tissue, including blood, thereby attenuating the light reaching the photodetector. Attenuation of the light through absorption by the tissue is linked to the volume of blood locally present in the tissue. Changes in local blood volume therefore alter the intensity of light reaching the photodetector.[3] [4] [5]

PPG sensors can be constructed to operate in either transmission mode or reflection mode, based on the configuration of the light source and the photodetector, as seen in Figure 2.1. In the transmission mode configuration, the light source and the photodetector are located opposite to one another, with the tissue placed between them. The light travels from the source, through the tissue, to the photodetector

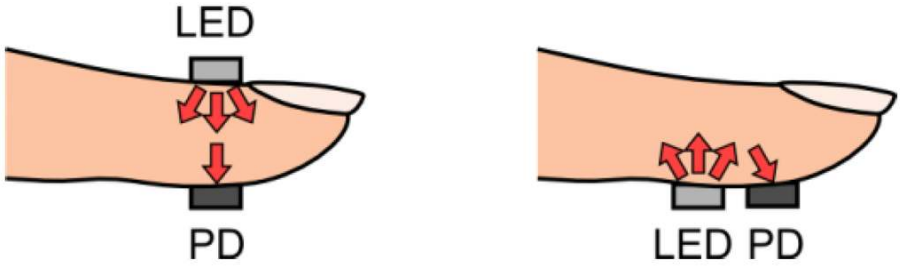

Figure 2.1: PPG sensors can be configured to operate in transmission (left) or reflection (right) modes, based on the placement of the light source and the photodetector. Reproduced from [6]under the Creative Commons Attribution 3.0 license (`https://creativecommons.org/licenses/by/3.0/`).

on the opposite side. Transmission mode is typically used for thinner tissues such as fingertips or earlobes. In the reflection mode configuration, the light source and the photodetector are placed adjacent to one another. Reflection mode is preferred for thicker tissues, such as wrists.[3] [4] [5]

PPG waveforms consist of both a varying 'AC', and a large, consistent 'DC' component, illustrated in Figure 2.2. The DC component accounts for the majority of the attenuation, which is due to the absorption of light by the skin, bone, venous blood, and other tissues. It is subject to low frequency variations due to factors such as respiration, vasoconstriction, Mayer waves (arterial pressure oscillations), and thermoregulation. Higher frequency fluctuations in the arterial blood volume are the source of the varying AC component. This was first described by Hertzman in the 1930s, as a source of error when measuring blood flow, occurring due to the presence of a large artery present beneath the sensor [7] [6]. These fluctuations are linked to the cardiac cycle, with the blood vessels containing more blood following systole, relative to the blood volume following diastole. Hence, the AC component can be used to determine a subject's pulse rate.[3] [4] [5]

The attenuation of light involves complex interactions with the tissue consisting of scattering, absorption and reflection. These have been shown to vary according

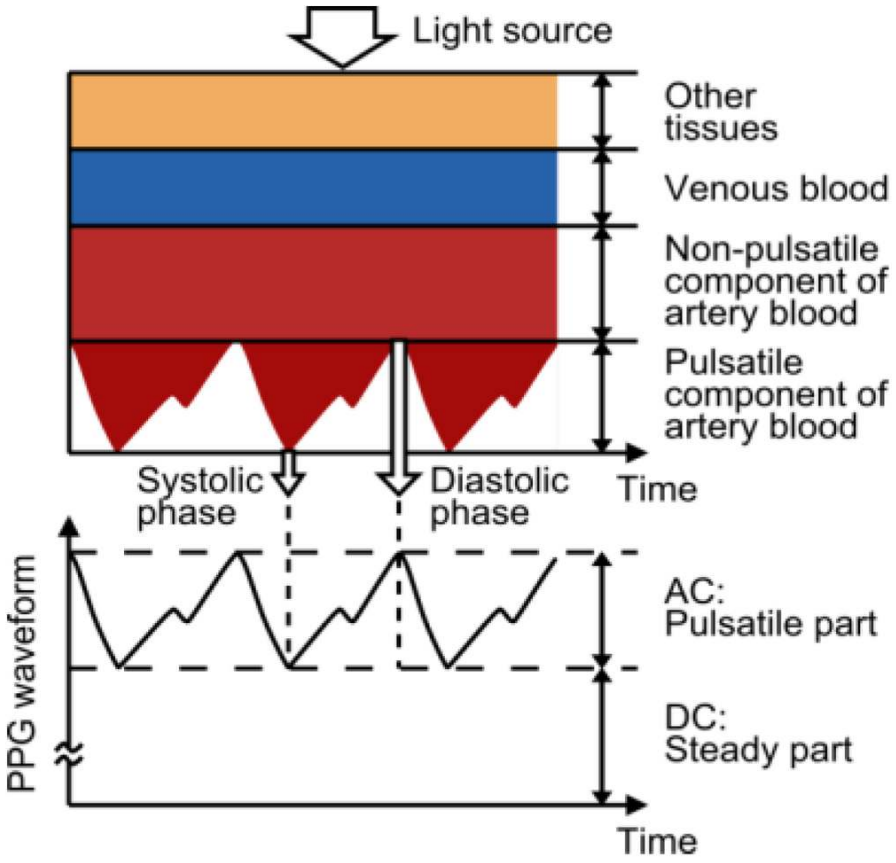

Figure 2.2: PPG waveforms consist of pulsatile (AC) and non-pulsatile (DC) components. The AC component is due to the differences in local blood volume following systole, relative to diastole. Reproduced from [6] under the Creative Commons Attribution 3.0 license (`https://creativecommons.org/licenses/by/3.0/`).

to the wavelength of light. Shorter wavelengths, such as 525nm (green), have higher attenuation through tissue than longer wavelengths, such as infrared light [8]. Thus, while green light PPG only accesses the vascular network in the skin, wavelengths such as red and infrared accesses deeper blood vessels [9]. Greater skin penetration depth of longer wavelengths was found to result in a more complex AC component which had a lower amplitude and was more susceptible to noise. Therefore, shorter wavelengths such as green light are preferred for wearables operating in reflective mode and are used in most commercial wristbands.[9] Skin pigmentation and the concentration of

melanin affect the absorption of the light by the tissue. While shorter wavelengths of light are readily absorbed by melanin [6], absorption is at its minimum in the 600-700nm (near red) range of the spectrum [4]. Despite this drawback, green light is preferred for commercial wearables, as it is considered to be less susceptible to corruption due to motion. As melanin is only present at the epidermis, in which blood vessels are not present, it was found to proportionally attenuate both the AC and DC components of the PPG [10]. Hence, it was proposed that the light intensity of the PPG could be increased for subjects with higher melanin without compromising the signal to noise ratio.[10] Other researchers have proposed methods using different wavelengths of light, such as the yellow-orange spectrum [11]. [4] [6]

2.1.1 Alternative Plethysmography Technologies

In 1989, Kamal et al [4] reviewed other techniques used for measuring blood flow at the skin level, including skin thermometry, thermal clearance, laser Doppler plethysmography, electrical impedance, and radio isotope clearance.

Skin thermometry is used to monitor changes in blood flow based on the skin surface temperature, relying on the blood flow used by the body for thermoregulation. However, this technique requires strict controls on the environmental temperature and humidity, which can be impractical in many clinical applications. [4]

Thermal clearance techniques use measurements of the difference between a heated component inside the probe and an unheated component which is in contact with the skin. Blood flow is assessed by analyzing the rate of heat dissipation away from the heated component caused by the blood flow in the skin. However, the technique requires a long lag time for a reading.

The laser Doppler method assesses blood flow using the Doppler shift caused by the velocity of the red blood cells in the blood. However, this requires accurate

mapping of the capillary network, which may not be practical. [4]

Radioactive clearance method measures the clearance of a radiopharmaceutical tracer, injected percutaneously. However, repeated measurements often cannot be conducted on the same subject due to limits on allowable radiation dose. [4]

2.2 Atrial Fibrillation

Atrial fibrillation (AF) is characterized by non-synchronous contraction of muscle fibres in the atrial chambers of the heart. The pumping activity of the atria are severely hampered, thereby preventing the atria from emptying their blood content to the ventricles. Furthermore, this affects the synchronicity of the electrical signal reaching the atrioventricular node, thereby causing arrhythmic ventricular contractions. Apart from reducing the efficiency of the heart, this also increases the risks of diseases and complications. AF has been associated with cognitive impairment, end-stage renal disease, venous thromboembolism, heart failure and various cardiac complications[12] [13]. The ineffectual pumping of the atria cause blood to pool in the chambers, resulting in blood clots that can travel to blood vessels causing stroke.[12] [14] There are three types of AF, classified based on temporal pattern [15]:

Paroxysmal: self-terminating without medical intervention

Persistent: continuing until medical intervention is taken

Permanent: continues despite medical interventions being attempted

Undetected (also called silent) AF is common and presents a severe risk of stroke and death. An estimated 20-30% of strokes occur due to AF, with an increasing number due to paroxysmal AF [16]. Furthermore, the AF itself can alter the electrophysiology of the atrial chambers of the heart which promote continued occurrences of AF [15]. Overall, AF causes a massive disease burden which affects the livelihoods

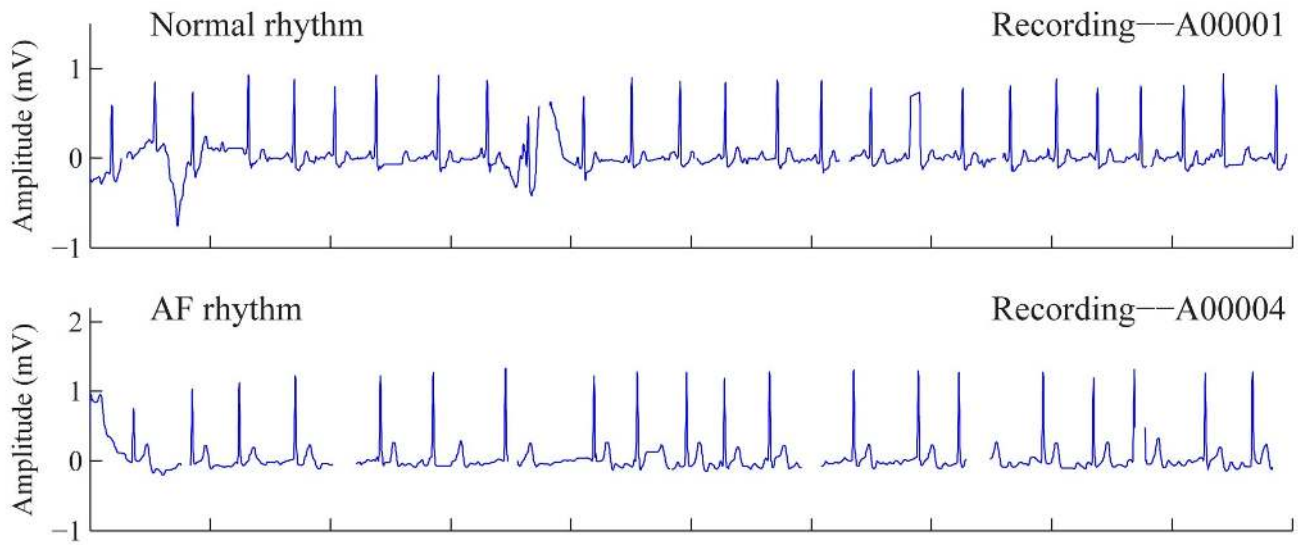

Figure 2.3: Electrocardiogram data for normal sinus rhythm and atrial fibrillation. The atrial fibrillation is distinguishable by the arrhythmic QRS complex occurrences and the lack of a P-wave. Reproduced from [18] under the PhysioNet Copying Policy.

of the patients and presents a cost to society. Thus, early detection is paramount to enable the introduction of medical interventions. The detection of paroxysmal AF is problematic as it is episodic in nature, thus detection of the condition at an early stage requires continuous cardiovascular monitoring. [17] The current technology being used for detection is an electrocardiogram Holter monitor, which is cumbersome for prolonged daily use due to the wires and chest electrodes required. A comparison between AF and normal sinus rhythm is shown in Figure 2.3 using electrocardiogram data. The atrial fibrillation is distinguishable by the arrhythmic QRS complex occurrences and the lack of a P-wave.

2.2.1 Photoplethysmography use in Atrial Fibrillation Detection

In 2015, Ferranti and Laureanti [14] proposed a method for detection of AF in patients using a PPG wristband. Their experiment was conducted with subjects remaining motionless and a short data collection time. The study was conducted on 70 subjects, from which 30 were diagnosed with AF, 31 were healthy, and another 9 were diagnosed with other arrhythmias. Subjects were asked to lie down on their back with the PPG wristband placed on their non-dominant arm for the 10-minute data collection procedure.

Noise detection was done through the analysis of the accelerometer signal (the accelerometer was part of the wristband), to identify intervals affected by motion artifacts. Intervals identified as being disturbed by noise were then eliminated, by replacing that interval with a flat zero signal. Data from a patient suspected of suffering from paroxysmal AF was also discarded from the study. [14]

Various features were extracted, including beat to beat intervals, morphology characteristics, time domain indexes, spectral characteristics, detection of multi-peak waves, and entropy. A correlation analysis was performed to eliminate redundant features, with principal component analysis being performed to identify the best combination of features. With the best features, a support vector machine (SVM) classifier was trained and evaluated using leave-one-out cross validation, providing an accuracy of 95.71% for distinguishing between AF, normal sinus rhythm, and other arrhythmias. [14]

While this method had a high accuracy, the subjects were required to remain in a steady state, and data with motion was discarded [14]. Thus, it is unlikely that this method can be directly applied to AF detection in an ambulatory, daily use setting.

2.3 Signal Quality Algorithms from Literature

Accurate cardiovascular parameters are difficult to derive when the PPG signal is corrupted by motion artifacts. Therefore, a reliable automated algorithm is required to classify the quality of PPG signal, so that noise-free and low-noise data can be isolated for used in cardiovascular analysis. This section presents three algorithms from the three major categories of PPG signal quality algorithms presented in literature.

2.3.1 Morphological Analysis

Sukor *et al* [19] proposed a method for the detection of noise-corrupted heart beats in the PPG signal using morphological features of the waveform.

Classification

Classification of the quality of the pulses was done through the use of a decision tree classifier (Figure 2.4). The first segment was for the classification of bad pulses from the other quality levels. This was based on three features; pulse amplitude, trough depth difference, and pulse width. The pulse amplitude is difference in height between the peak of the pulse and its preceding trough. Trough depth difference is the height difference between two successive troughs. Pulse width is the time difference between two successive troughs. Thresholds are used for the classification of bad pulses based on these features.

An average pulse template is constructed by aligning the pulses from a recording at the pulse peaks, then obtaining the means pulse shape. Two features are used to distinguish poor pulses and good pulses from the remaining data; Euclidean distance and amplitude ratio. The Euclidean distance is taken between the current pulse and the mean pulse shape. The amplitude ratio is the difference in pulse amplitude be-

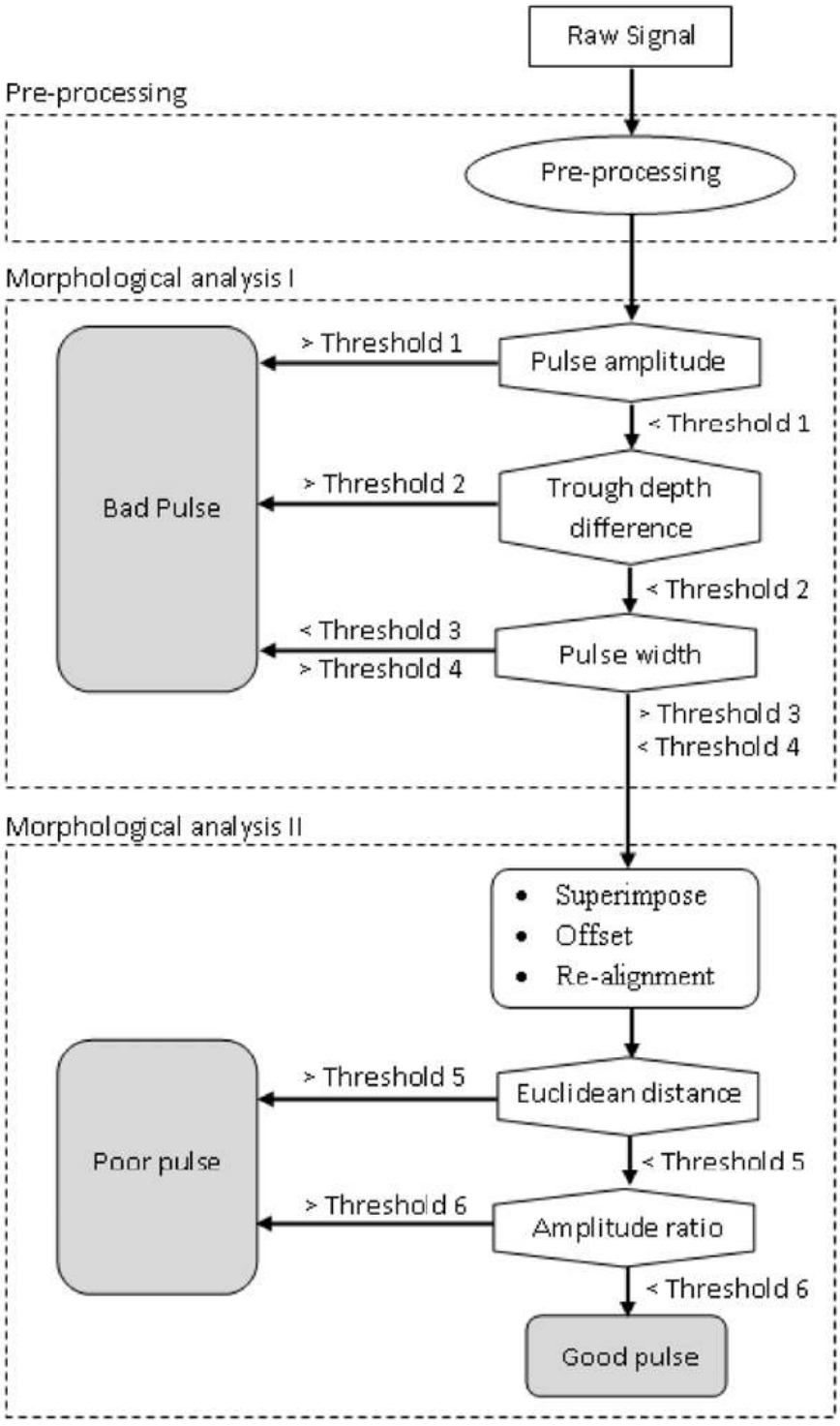

Figure 2.4: Classification of signal quality to good, poor and bad pulses using decision tree. Reproduced from [19].

tween the current pulse and the mean pulse shape. Thresholds are used to distinguish the two quality levels in the decision tree.

Dataset

Validation of the algorithm is done with a dataset of 104 recordings acquired from 13 healthy subjects. PPG and ECG are simultaneously recorded using a reflective mode finger-clip pulse oximeter sensor and a Lead II configuration ECG, respectively. Subjects are asked to perform eight different hand movements to simulate a variety of motion artifacts, with a separate recording per hand movement. Each recording is approximately 60 seconds, with 20 seconds in the centre of the recording allocated for the movement. The hand remained stationary, resting on a table top, for the first and last 20 seconds of the recording.

Gold Standard

The PPG and ECG signals were subject to linear detrending, followed by zero-phase band-pass filtering at 0.5-5Hz and 0.5-150Hz, respectively. Quality levels were defined as good, poor and bad. Good pulses were defined as those having a standard PPG morphology and widths and amplitudes similar to neighbouring good pulses. Poor pulses were defined as pulses with widths similar to those of good pulses, but with differing amplitudes and morphologies. Bad pulses are defined as those with significantly different widths, amplitudes, and/or morphologies compared to good pulses.

Ratings for the quality of pulses were assigned by two raters independently. Pulses for which there was a difference in rating were re-assessed by both experts together to present a reconciled rating.

Discussion

This yielded an accuracy of 83%, with a sensitivity and specificity of 89% and 77%, respectively. While the hand movements performed during the data collection in study may have been representative of motions occurring during ambulatory daily use, the dataset was collected from a finger-clip PPG sensor and may not be consistent with motion artifacts present in wrist-based PPG systems.

2.3.2 Dynamic Time Warping based Template Matching

Due to the non-stationary nature of pulsatile blood flow, caused by changes in heart rate, cardiac output and sensor-location variability, the use of simple beat to template matching approaches for signal quality assessment of PPG can be challenging. To address this challenge, Li and Clifford [20] proposed a signal quality algorithm utilizing dynamic time warping (DTW) to stretch each beat to determine the best match to the template. The DTW and the other features are used with a multi-layer perceptron neural network to classify the signal quality.

Classification

Two classification methods are used to determine the signal quality from the features; a simple heuristic method and a multi-layer perceptron (MLP) based machine learning system. The heuristic method combines the four features into a single metric, the fused signal quality index (qSQI), based on the values of the four features. The qSQI is calculated for each beat segment, and a mean qSQI is computed for each segment by averaging the qSQIs from each beat within the 6-second segment. If the mean qSQI is greater than a threshold value, then the segment is classified as good, otherwise it is classified as bad. The threshold value is varied between 0 and 1 in steps of 0.01, and

the accuracy computed at each point. The threshold value with the greatest accuracy is then used for the classification of the test.

For the machine learning system, two MLP based systems were used; the first system with four features, and the second system with 6 features. The four-feature system was trained using the aforementioned SQIs. The six-feature system used the qSQI and the number of beats detected within the 6s segment as features, in addition to the aforementioned SQI features.

Dataset

A PhysioNet dataset taken from 104 critical care recordings was used to evaluate the algorithm. The data set was annotated with occurrences of asystole and ventricular tachycardia arrhythmias. Data segments with a length of 6 seconds were selected beginning 5 seconds prior to the arrhythmia event.

Gold Standard

The quality of PPG segments was rated by two raters as good, bad or uncertain, with disagreement between raters arbitrated by a third rater. Data segments with a rating of uncertain were discarded, leaving of 1055 PPG segments, 825 good and 230 bad, to be used for the evaluation of the algorithm. It should be noted that this did not use wrist-based PPG. Individual beats were denoted using the *wabp.c* Arterial Blood Pressure beat detector from PhysioNet, and fiducial marks used to segment each individual beat. A template for PPG beats was generated by averaging all the beats in each of the 30 seconds of data, to be used as a basis for comparison.

Features

Four features are extracted for the algorithm; direct matching SQI (Signal quality Index), linear resampling SQI, dynamic time warping SQI, and clipping detection SQI. A correlation coefficient is calculated for each 30 second window between each individual beat and the beat template. The beat is selected from the fiducial mark to the length of the template beat. Any negative correlation values are zeroed, thus providing a direct matching SQI between 0 and 1.

For the second feature, the linear resampling SQI, each individual beat is linearly stretched or compressed to match the length of the template. A correlation coefficient between the linearly resampled beat and the template is computed, and negative values are zeroed.

The dynamic time warping SQI was calculated using a similar procedure, with DTW used to resample the beat. DTW provides a non-linear time-based approach to stretching the beat. Supposing there are two time series T (template) and B (beat), with lengths n and m, respectively. An n x m matrix is constructed, with each element corresponding to the distance a point on T and a point on B. In the matrix, an optimal path from index (0,0) to (n,m) must be found to reduce the total distance of each index on the path. An example of this path is illustrated in Figure 2.5 b, by the black squares in the matrix, and the resulting alignment between the template and the beat is shown in Figure 2.5 c. With the aligned time series, a correlation coefficient is taken, with negative values being zeroed.

The fourth feature, the clipping detection SQI, hysteresis thresholding was used to identify periods of saturation to a maximum or a minimum value, defined as being clipped. The clipping detection SQI is the percentage of the beat that is not clipped.

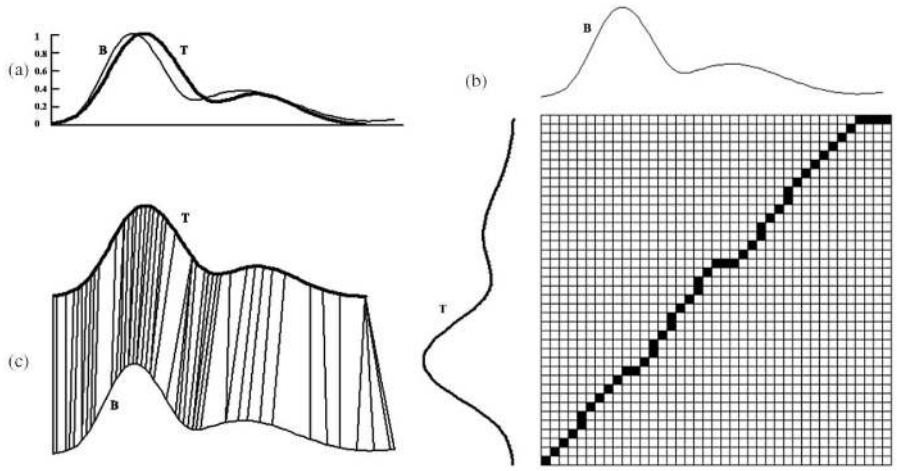

Figure 2.5: An example of the process of Dynamic Time Warping is shown. The template and the current beat are matched (a), then a matrix of the distances between specific points is constructed, in which the black squares show the path with the lowest cumulative distance (b). (c) shows the alignment of points in the template and the current beat. Reproduced from [20].

Discussion

A test performance with an accuracy of 95.2% was achieved with the six-feature MLP system, which was the highest accuracy obtained from the test. While the algorithm provides high accuracy, the data set on which it was tested is not from wrist-based PPG. Furthermore, the dataset was not collected from an ambulatory, daily use scenario, but rather from critical care patients experiencing cardiac events. It ought to be noted that the study only used two quality levels, and discarded PPG segments whose quality could not be clearly defined as good or bad, thus potentially leaving out numerous mid-level segments from the study.

2.3.3 Fusion Approach

Traditional template matching quality algorithms require accurate beat segmentation to isolate individual heart beats for matching. However, the presence of noise

corruption in the PPG signal, such as that found in ambulatory settings, renders the segmentation process more prone to error. Wander and Morris [21] hypothesized that quality assessment techniques dependent upon accurate segmenting would perform poorly in the presence of periodic motion artifacts, compared to non-segmenting techniques. Thus, they proposed an algorithm utilizing both template matching (requiring segmentation) and morphological features (not requiring segmentation) to outperform either technique used alone.

Classification

Using the extracted features, a classification and regression tree (CART) is trained and tested using leave-one-out cross validation, in which the entire data from a subject was left out at each of the 11 iterations. The algorithm performs with a Pearson correlation of 0.9263 and a mean square error of 0.4627.

Data

The data consisted of signals collected using wrist-based reflective PPG, a 3-axis accelerometer at the same location as the PPG sensor, and a chest strap heart rate monitor. While wearing the sensors, participants were asked to perform three behavioural tasks; standing still, walking in place, and jogging in place. After an initial recording of 30 seconds of standing still, the three behavioural tasks were performed three times, presented in a randomized order, lasting for 80 seconds each, thus resulting in a total recording of 12 minutes per subject. The data was collected from 11 participants, all with no known history of cardiovascular illness.

Gold Standard

The signal quality level was classified by raters on a scale of 1 to 4. The definition of the quality levels are as follow:

4) All beats easily visually identifiable

3) More than half of the beats are easily visually identifiable

2)Fewer than half, but more than one beat is visually identifiable

1) One beat or none are visually identifiable

The data was split into non-overlapping 7 second segments. Three expert raters were used to visually annotate the signal quality of the signal, with two raters rating every alternate segment, resulting in two ratings per segment. Along with the PPG data, the raters were also presented with the corresponding accelerometer data to reduce the probability of mistaking periodic noise for legitimate PPG pulses. The final manual quality rating (MQR) assigned to each segment was the mean of the rating provided by the two raters.

Features

Prior to feature extraction, the PPG signal is put through a zero-phase, 4^{th} order Butterworth band-pass filter with poles at 0.5Hz and 50Hz.

Direct Signal Statistics

Kurtosis: The kurtosis was taken for each segment, then log transformed to reduce the impact of outliers.

Autocorrelation peak strength: Correlations of the signal with itself at various lags was computed, forming a correlogram. This is smoothed by a rectangular filter 0.05 seconds wide. Then the amplitudes of the peaks in the correlogram were computed,

with the largest amplitude being used as the feature value for the segment.

Spectral power: Using the Welch method of spectral density estimation with a window length of 1024 samples and 50% overlap, the spectral power at frequencies of 1, 3, 5, 7, 9, 13, 17, 21, 25, and 29Hz were used as features for each segment.

Template Match Statistics

Segmentation of the individual beats in each PPG segment was done using the method of repeated Gaussian filters. Three features are extracted during the segmentation process:

Autocorrelation HR score: maximum peak to trough difference of the autocorrelation function of the data window

Gaussian correlation score: correlation coefficient of the repeated Gaussian filters with the derivative of the PPG signal

HR estimate score: product of the autocorrelation HR score, the Gaussian correlation score, and the heart rate penalty from the segmentation procedure

An initial template set was created by randomly selecting 20 high quality beats from the initial 30 seconds of the data collection in which subjects were standing still. The template is updated for each subject to converge to a subject-specific template. Beats constituting the template are incrementally replaced during the matching process by beats with high correlation scores to the template. Features extracted from the matching process were; mean beat correlation, median beat correlation, minimum beat correlation, and maximum beat correlation.

Sample to Sample Transition Statistics

The transition statistics are computed to utilize the quasi-periodic nature of the PPG waveform. Transition statistics study the change between subsequent data points

(samples). For each segment, the data is normalized to an interval from 0 to 1 and quantized to 16 discrete values. A transition matrix (illustrated in Figure 2.6) is constructed for sample n versus sample n+1, which is then normalized to provide the probability distribution. The differences in transition matrix by quality are shown in Figure 2.6. Feature reduction and fusion is performed to deal with the high dimensionality of the transition matrix, as each element in the matrix is a feature. Subsequently, a support vector machine (SVM) is trained and subsequently used to provide rating on the aforementioned 1 to 4 scale.

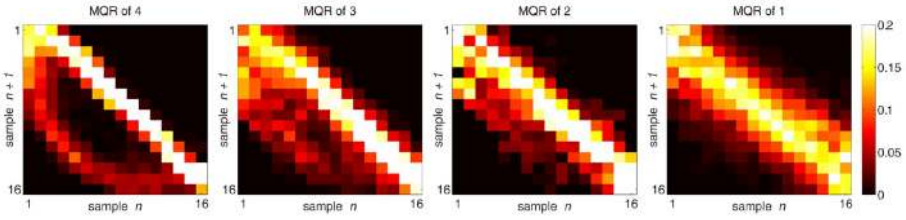

Figure 2.6: Transition Matrix for each signal quality level. The matrix consists of sample n versus sample n+1 . Reproduced from [21].

Discussion

While this technique was validated on wrist-based PPG, the data was collected by inducing periodic motion artifacts by asking subjects to perform specific behavioural tasks. The data used was not collected through ambulatory, daily use, which would be expected to contain more non-periodic noise.

2.3.4 Discussion

While these algorithms provide a good preliminary foundation, none of the algorithms were developed using ambulatory data with participants conducting their daily activities. Motion artifacts were induced by asking participants to performs various tasks,

which do not necessarily cover the wide variety actions and movements performed by individuals in an ambulatory environment. Furthermore, while visual assessment is used as the gold standard for quality, studies of the reliability of the gold standard being used to develop and evaluate the algorithms are lacking.

Chapter 3

Data and Devices

Simultaneous wrist-based photoplethysmography (PPG) and electrocardiogram (ECG) data was collected from 26 healthy participants for a period of 24 hours as they performed their daily routine. The study was approved by the Carleton University Research Ethics Board.

3.1 Devices

3.1.1 Photoplethysmography Wristband

The Empatica E4 wristband, shown in Figure 3.1, was used to obtain the PPG signal from the wrist. The Empatica E4 is a wrist-based device, similar to a smartwatch, equipped with other sensors such as a temperature sensor (infrared thermopile), a 3-axis accelerometer, and a skin conductance sensor. These signals are recorded continuously and stored internally on the device. The Empatica E4 was chosen for use in the study as it allowed easy access to the PPG waveform. Other smartwatches and fitness wristbands containing a PPG sensor were considered, however, these did not provide a straightforward manner for accessing the PPG waveform.

The PPG and accelerometer signals were used in this study. The PPG signal is sampled at a rate of 64Hz, while the 3-axis accelerometer signal is sampled at a rate of 32Hz [22]. The device has sufficient internal memory for 48 hours of continuous recording. In order to determine the battery life, the device was left recording until shutdown; the device was tested to approximately 48 hours. However, battery life is expected to decrease over time, with the technical specifications sheet for the device [22] stating a battery life of 36 hours. Thus, the device meets the necessary requirements to be used for continuous 24-hour data collection desired for this study. The device is splash resistant, thus hand washing was not restricted for participants, however, participants were asked not to submerge the device at any point.

Figure 3.1: The Empatica E4 wristband was the source of the PPG Signal.

The PPG sensor of the Empatica E4 consists of four LEDs; two using light in the red wavelength spectrum, and two in the green. These are located at the bottom

Figure 3.2: The Empatica E4 wristband's PPG sensor consists of four LEDs and two photodiodes.

of the device, shown in Figure 3.2, facing the posterior side of the wrist. The LEDs of each colour are located diagonally opposite to each other, with two photodiodes located between them. A barrier separates the photodiodes from the LEDs, thus reducing the effects of cross-contamination. According to vendor, the green signal is expected to contain the pulse wave information, while the signal obtained from the red wavelength is used as a reference light measurement to remove noise interference [23]. This is likely due to varying attenuation levels of the two wavelengths as they pass through tissue, as discussed in Chapter 2.

The raw signals measured by the device are not made available. Instead, the device uses a proprietary algorithm, internally combining the signals obtained from the green and red wavelengths to produce the PPG pulse signal. The device attaches to a base

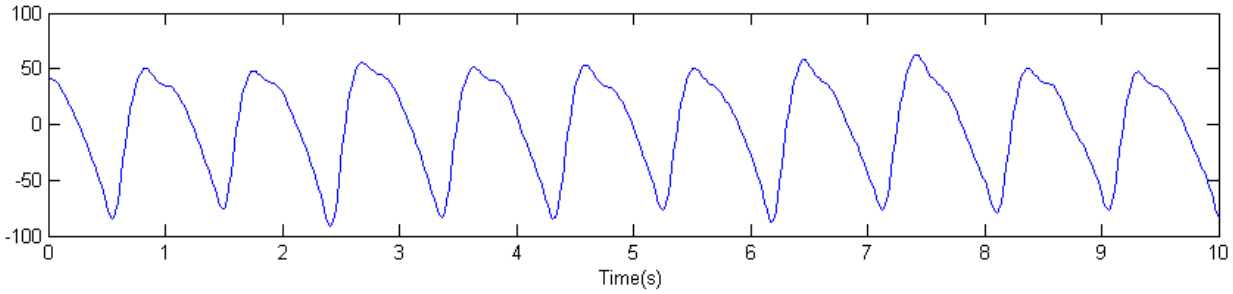

Figure 3.3: Example of PPG Signal from the Empatica E4.

with a micro USB connector, which enables battery recharging, and downloading of the data onto a computer. The data is accessible in *.csv* format, with the recording timestamps in UNIX time format. An example of data collected from the Empatica E4 is shown in Figure 3.3.

3.1.2 Electrocardiograph Holter Monitor

The General Electric Seer Light Extend Holter monitor, was used to obtain the ECG signal from the thorax. The Holter monitors were obtained from the University of Ottawa Heart Institute (UOHI).

Figure 3.4: The General Electric Seer Light Extend Holter monitor was the source of the ECG Signal.

The device consists of two parts; the main Holter case where the data is recorded, shown in Figure 3.4, and the cable linking the case to the electrodes. The electrodes from the device are attached to various sites on the thorax of the participants. As the device is not resistant to water damage, participants were asked to modify their daily routine to avoid exposing the Holter to water.

The Holter has a sampling rate of 125Hz with 10-bit resolution, at which it can record 48 hours of data [24]. The device records ECG in three leads; modified aVF, modified V1, and modified V5 [25], all of which were recorded simultaneously.

3.2 Participants

Participants were recruited from two broad demographics; 1) elderly, defined as individuals 65 years of age or older, and 2) non-elderly, defined as anyone not included in the elderly group. This enables the evaluation of differences in the quality of wrist-based PPG between the two groups, as the elderly group is expected to have different types of activities in their daily schedules.

Healthy mobile subjects were recruited for participation in this study. Mobility was defined as the ability to walk without requiring assistive devices. An inclusion criterion of mobility was used for the study, as the motion artifacts due to movement during daily activities were expected to be the primary source of signal artifacts. The daily routines of participants with compromised mobility would be expected to include considerably less motion than the average person. Thus, the inclusion of participants with compromised mobility would bias the dataset, producing data with uncharacteristically high signal quality.

The elderly participants were recruited through an advertisement in the Age Well Solutions Newsletter, and through contacts with other researchers. The non-elderly

participants were recruited from Carleton University students and researchers. The study was conducted with 26 participants; 10 participants from the elderly category, and 16 participants from the non-elderly category. Further demographic information such as age, gender and ethnicity were not collected, as per the Research Ethics application for the study. To the best of our knowledge, these statistics were not regarded as being relevant to the signal quality of the devices. Skin tone, specifically skin melanin concentration, is expected to affect the PPG signal. However, this information was also not collected as there was no practical method to quantify the melanin concentration for the study. Additionally, it was not known whether the Empatica E4 automatically adjusted the light intensity of the LEDs to improve signal-to-noise ratio, which is one of the recommended methods to compensate for users with high concentrations of melanin.

3.3 Experimental Setup

Participants were asked to wear the Empatica E4 wristband and the GE Seer Light Extend Holter monitor. Skin preparation was done for the Holter monitor at each of the seven electrode sites. Abrasive skin preparation gel, Nuprep, was applied to each electrode site with a Q-tip, then removed with a gauze pad. Rubbing alcohol was then used to clean each electrode site, after which the electrodes were attached. Kendall foam electrodes, recommended by the Holter monitoring group at the UOHI, were used for this experiment.

Wires from the cable were attached to each electrode, corresponding to the electrode positions shown in Figure 3.5. To strengthen the adherence of the electrodes to the thorax, medical tape was applied to each electrode head. Stress loops were made with loose wires, and attached to the thorax using medical tape. This was done

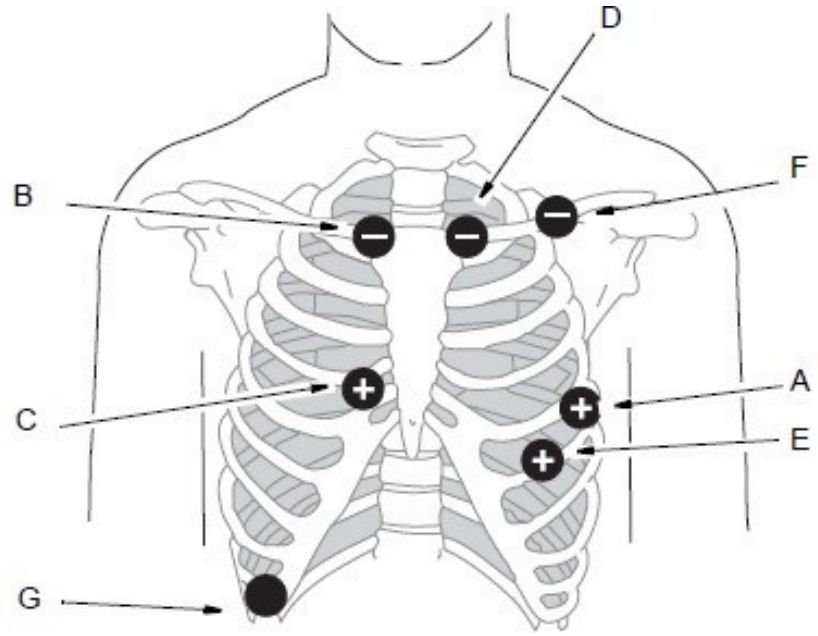

Figure 3.5: Diagram depicting the electrode placement for the ECG Holter. Reproduced from [25].

to prevent detachment of the electrodes during motion, by reducing tension at the electrodes caused by pulling of the wire.

As per the recommendation of the Holter monitoring group at the UOHI, for female participants, the electrode corresponding to site C in Figure 3.5 was moved to the centre of the thorax, onto the sternum.

The Empatica E4 wristband was affixed with the sensor facing the posterior side of the participant's wrist on their non-dominant arm. It was worn similarly to a wristwatch, as shown in Figure 3.1. The non-dominant arm was expected to engage in less movement relative to the dominant arm, thus providing the potential for superior signal quality.

For a period of 24 hours, the participants wore both devices for simultaneous monitoring. Participants were required to refrain from any activities that may damage the devices. As the Holter monitor was susceptible to water damage, participants

were asked not to swim or bathe during the 24-hour monitoring period. With this exception, they were asked to continue with their daily routine as much as possible.

After the data collection period, the devices were removed and the data from the devices were extracted. The data from the Empatica E4 was extracted through the Empatica manager software. The Holter was taken to the UOHI, as it required the GE MARS Utility software for data extraction. The devices were then cleansed with alcohol prior to use for the next participant.

3.3.1 Limitations of Experimental Setup

Certain limitations were present in the experimental procedure owing to the practicality of scheduling with participants. An attempt was made to ensure that data collection would be conducted for a full 24-hour period. However, occasionally, participants were unable to meet with the researchers at the appropriate times due to their daily schedules, thus, the data collection period was sometimes slightly shorter than 24 hours.

In other cases, participants were unable to meet with the researchers the day following the device setup, and were therefore instructed to remove and power off the devices themselves. However, occasionally, participants were unable to power off the devices, or forget how to power off the devices. In such cases, the devices kept recording until they were retrieved by the researchers. In these cases, the data after the first 24 hours was not used for analysis.

The Empatica E4 wristband was worn as tightly as the participants felt was comfortable. Since the tightness was based on acceptability by the participant, some participants wore the devices more loosely than others. Those that wore the device more loosely are likely to provide inferior signal quality, as the wristband had a greater freedom of movement, and motion is expected to be a cause of noise corruption in

the signal.

Participants were not required to keep an activity log for the study, therefore, their sleep and exercise times are not known. This information may have enabled additional analysis and understanding of the data, thus, and activity log ought to be included in future studies.

Chapter 4

Development of the Gold Standard

Establishment of the gold standard was done by selecting a subset of data and requesting manual classification by raters based on visual assessment. This was necessary as there is no established gold standard used for wrist-based PPG quality classification.

Visual assessment can be subjective to the perspectives of the raters. Previous publications used gold standards established by a very small number of raters [21] [20] [19], hence the extent of the subjectivity is unknown. To better understand the variability of ratings due to subjectivity of raters, our gold standard was created using a compilation of ratings from 17 raters. Statistics were then computed to quantify the extent of quality classification agreement between raters.

4.1 Selection of Data

The entire PPG data set for each participant was subdivided into 10-second segments. The length of 10 seconds was chosen to maintain consistency with previous work [2]. A random number generator in Matlab was used to select 39 of the segments for analysis, from each of the 26 participants. Thus, a set of 1014 non-overlapping, non-

continuous PPG segments was compiled to be used for analysis. As the segments were chosen randomly, there were no controls in place to ensure equal class representation in the selected set of segments. As this set is going to be used to create the ground truth, there was no pre-existing classifications that could be used to ensure equal class representation in the set. Manual selection by the researchers would have resulted in bias towards their considerations of the classifications. Thus, a random selection process was chosen. It is expected that the random selection process would result in a class representation roughly proportional to the class representation in the overall data.

4.2 Definition of Quality Levels

Signal quality levels were chosen from 1 to 5 for rating the quality of the PPG signal for each 10-second segment, with 1 representing the lowest quality, and 5 representing the highest quality. The quality levels were defined by the percentage of the segment for which clear pulses with discernible peaks were identifiable. A class 5 segment must have all identifiable pulses for the entire data segment, class 4 for at least 75% of the segment. This goes down to class 1, for which less than 25% of the segment has identifiable pulses. This is shown in Figures 4.1 - 4.5. This criteria was based on the quality level definitions used by Wander and Morris [21] in their publication, which used the number of peaks visible within the data segment.

4.3 Graphical User Interface

A graphical user interface (GUI) was created in Matlab for the raters to annotate the PPG segments. This is shown in Figure 4.6. The 17 raters were selected from among

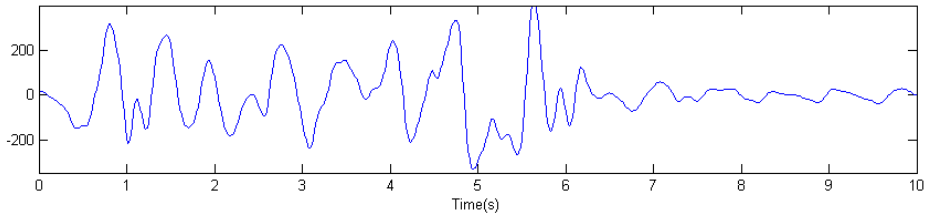

Figure 4.1: Example of PPG Signal belonging to Class 1.

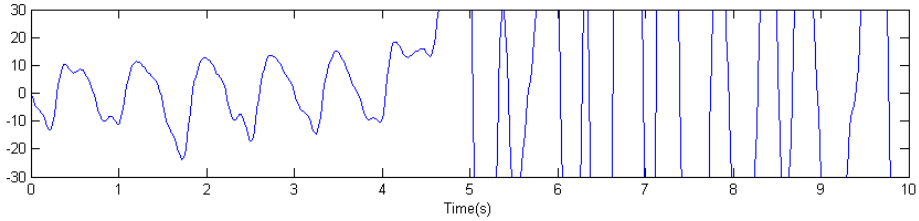

Figure 4.2: Example of PPG Signal belonging to Class 2.

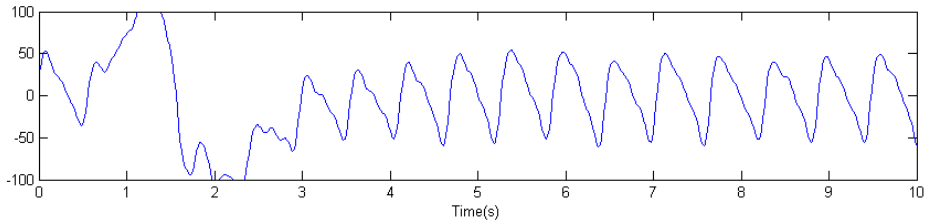

Figure 4.3: Example of PPG Signal belonging to Class 3.

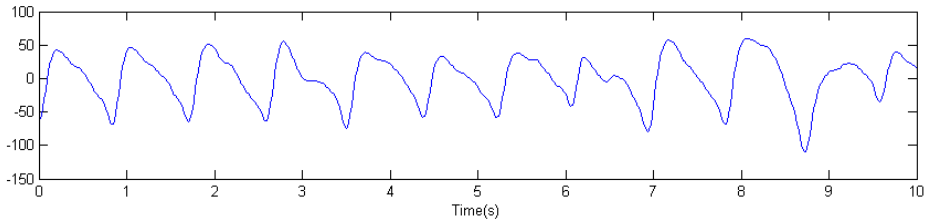

Figure 4.4: Example of PPG Signal belonging to Class 4.

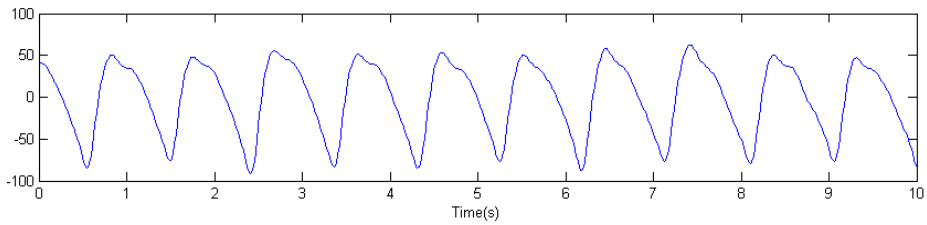

Figure 4.5: Example of PPG Signal belonging to Class 5.

biomedical engineering students at Carleton University. Each rater was provided with the definition of the quality classes, as well as two examples of each class, as identified by the researchers. The raters were not required to annotate the entire set of segments in one sitting, as the GUI saved the annotations after each segment. This enabled them to close the application, and then reopen it when they were ready to resume. Raters were provided with a $10 gift card to participant in the annotation process. This process was approved by the Carleton University Research Ethics Board.

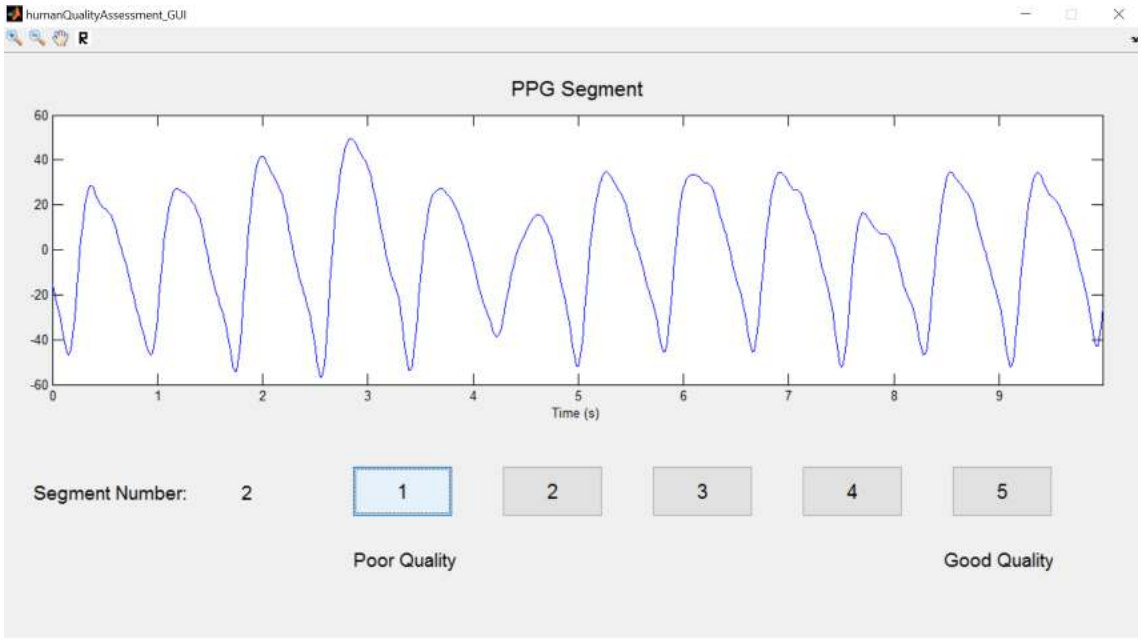

Figure 4.6: The Matlab graphical user interface used by the raters to annotate the PPG data segments.

4.4 Agreement between Raters

Annotations made by the raters were compiled into a table. The table depicts the number of raters who annotated each quality class for each PPG segment. Each row represents a PPG segment, while each column represents a quality class. For each PPG segment, the class chosen by a plurality of raters was used as the actual class

for the segment. Therefore, the gold standard was taken as the mode of each row in the table.

Agreement between raters, P_i, was computed for each PPG segment using Equation 4.1, as provided Fleiss [26]. According to this measure, a P_i value of 1 indicates that all raters agreed on the classification of the segment. The P_i values from all the PPG segments were collated together to produce the histogram in Figure 4.7. This figure shows the number of segments over the range of P_i values.

$$P_i = \frac{1}{n(n-1)}\Big(\sum_{j=1}^{k} n_{ij} - n\Big) \tag{4.1}$$

where n is the total number of raters, k is the number of quality classes, i is the segment number, and n_{ij} is the number of raters who classified segment i to class j.

Only approximately 11% of the PPG segments had perfect agreement between all 17 raters. This indicates that the classification of the majority of PPG segments had an element of subjectivity, despite a clear set of rules governing class membership. Approximately 58% of the segments had a P_i greater than 0.5, with the remaining 42% showing poor agreement between raters.

To obtain a normalized measure of the strength of agreement between raters, Fleiss' Kappa, κ, was computed.The agreements (P_i) computed for each segment using were averaged to obtain $\bar{P}$, the formula for which is shown in Equation 4.2. This value is referred as the overall extent of agreement.[26]

$$\bar{P} = \frac{1}{N}\Big(\sum_{i=1}^{N} P_i\Big) \tag{4.2}$$

where N is the total number of segments, and P_i was computed from Equation 4.1.

The proportion of all ratings assigned to each class, P_j, was computed using

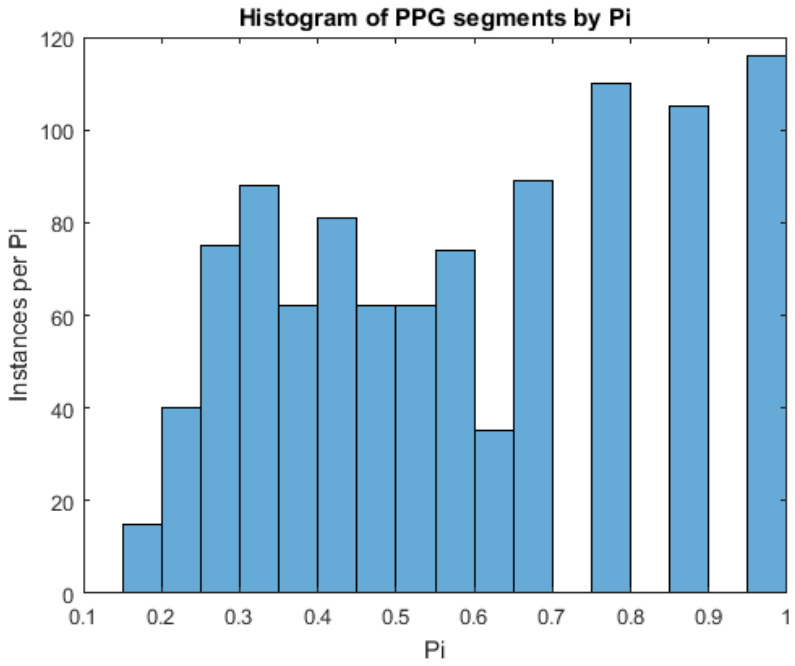

Figure 4.7: Histogram showing the distribution of PPG segments used for analysis, organized by the agreement between raters per segment, P_i.

Equation 4.3

$$P_j = \frac{1}{Nn}\Big(\sum_{i=1}^{N} n_{ij}\Big) \tag{4.3}$$

where N is the total number of segments and n_ij is the number of raters who classified segment i to class j.

According to the Fleiss method [26], a certain degree of agreement between raters is expected due to chance. Using the obtained P_j values, the mean proportion of agreement expected if ratings were made randomly, $\bar{P}_e$, was calculated using Equation 4.4.

$$\bar{P_e} = (\sum_{j=1}^{k} P_j^2) \tag{4.4}$$

where P_j was obtained from Equation 4.3.

The total possible extent of agreement beyond the agreement due to random

chance is represented by 1-$\bar{P}_e$. The agreement obtained in excess to random chance is represented by $\bar{P}$-$\bar{P}_e$. These two measures are used to compute Fleiss's Kappa, κ, as shown in Equation 4.5

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \tag{4.5}$$

A Fleiss' Kappa of $\kappa = 0.4605$ was obtained for this study. According to the benchmarks established by Landis and Koch [27] shown in Table 4.1, this indicates that overall, there was moderate agreement between raters, and that the agreement was not due to chance.

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Table 4.1: Interpretation of κ ranges. Reproduced from [27].

4.4.1 Rater Agreement by Class

Further histograms were constructed to depict the variation in rater agreement for each quality class. These are shown in Figures 4.8 - 4.12. High agreement among raters is seen in most segments classified as classes 1 and 5, seen in Figures 4.8 and 4.12. However, for classes 2, 3, and 4, the majority of segments had a P_i less than 0.41, indicating that the raters had a lower extent of agreement for segments belonging to those classes, compared to classes 1 and 5. However, even for classes 2-4, the strength of the agreement was greater than agreement expected due to random chance, indicating that there was a trend for raters to select those classes.

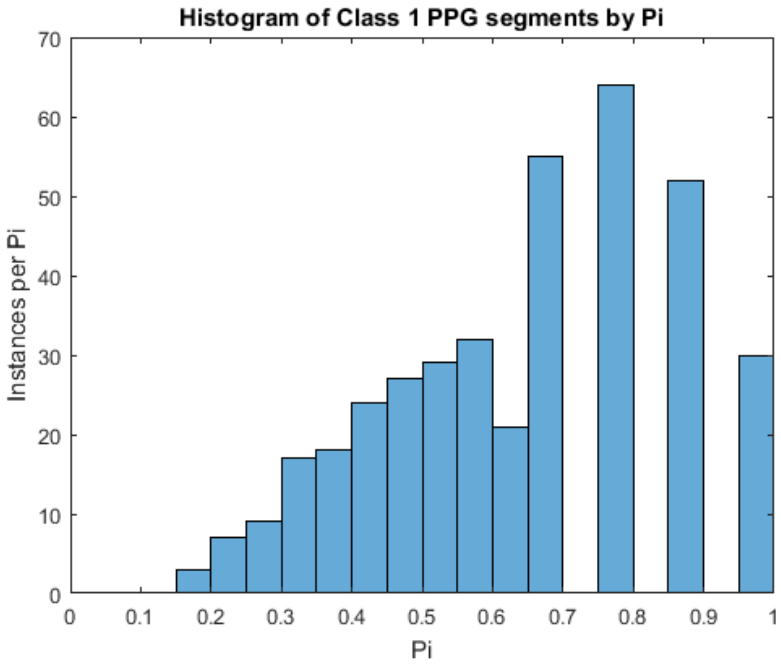

Figure 4.8: Histogram representing the distribution of agreement strengths, P_i, for segments classified as Class 1.

Overall, the extreme cases of high quality (class 5) and low quality (class 1) had the best inter-rater agreement, while the mid-level classes had much more disagreement. In segments with high levels of motion, the PPG signal is corrupted to the extent that no pulses are visible, enabling classification of such segments without much dispute between raters. Similarly segments which are noise-free were classified without much disagreement. Much of the disagreement with classes 2-4 was likely over what constituted a pulse with a clear peak, which is subjective in certain cases.

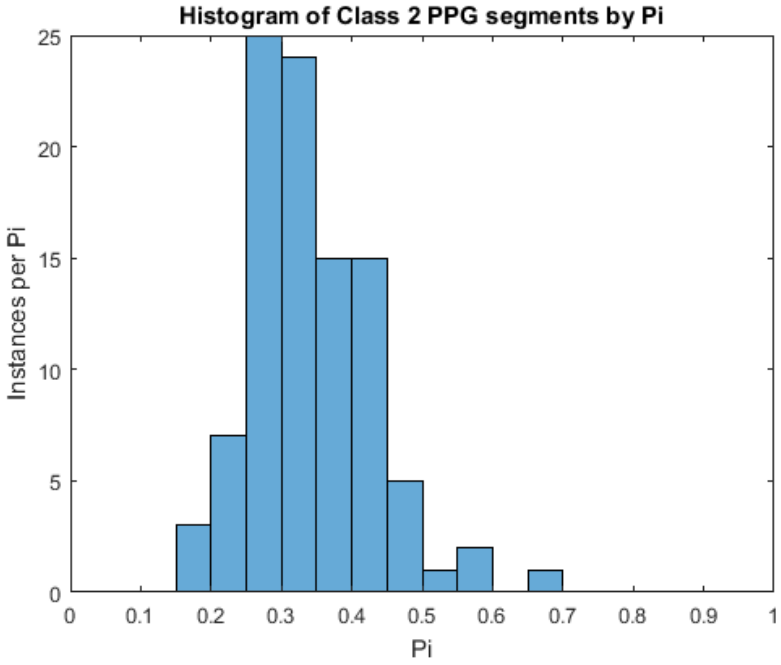

Figure 4.9: Histogram representing the distribution of agreement strengths, P_i, for segments classified as Class 2.

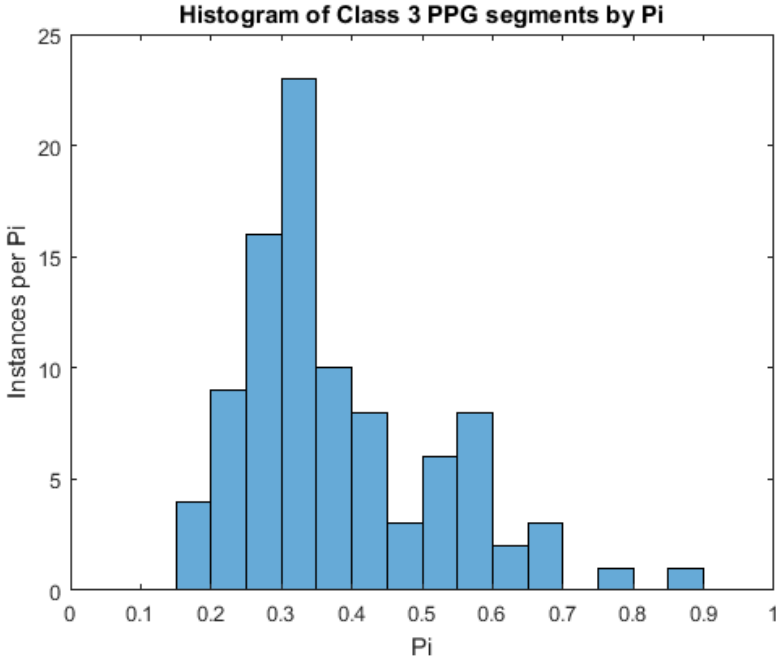

Figure 4.10: Histogram representing the distribution of agreement strengths, P_i, for segments classified as Class 3.

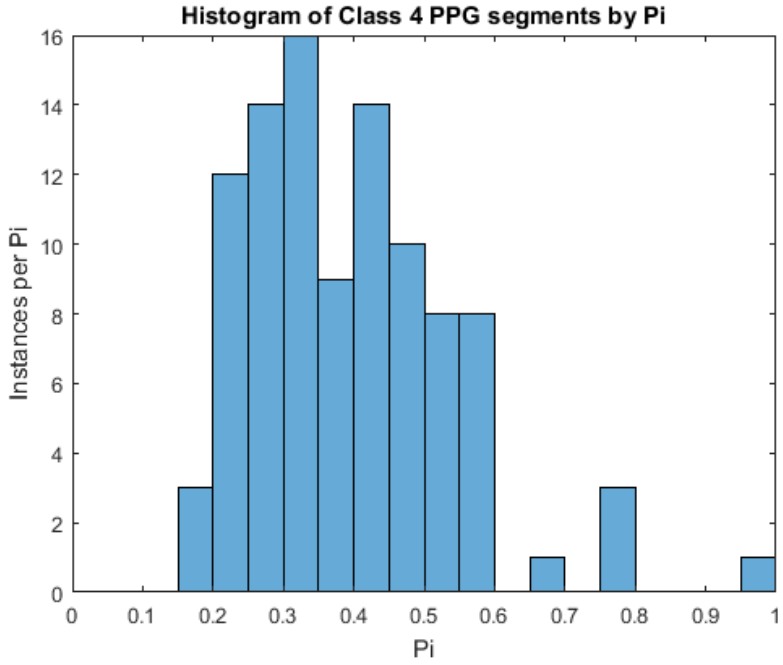

Figure 4.11: Histogram representing the distribution of agreement strengths, P_i, for segments classified as Class 4.

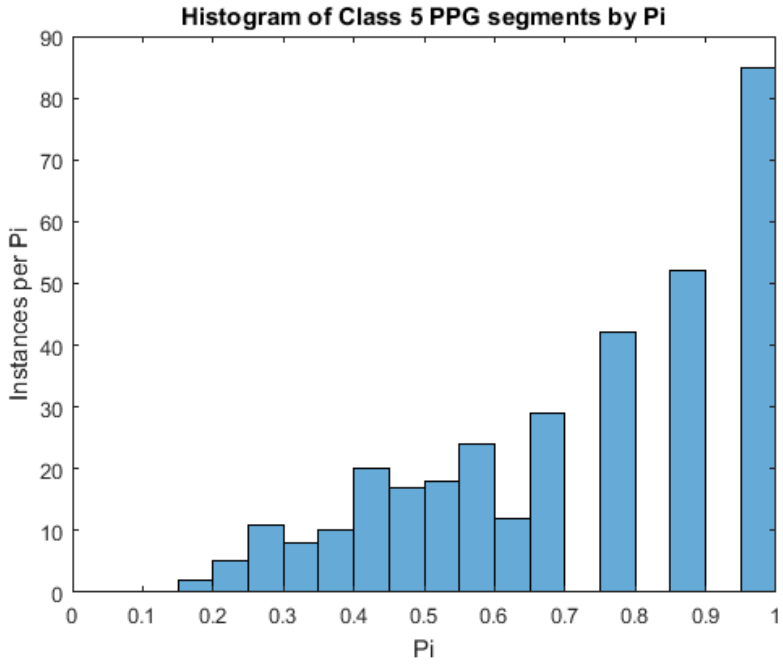

Figure 4.12: Histogram representing the distribution of agreement strengths, P_i, for segments classified as Class 5.

Chapter 5

Evaluation of Features

Features were extracted from each of the 1014 ten second data segments used in Chapter 4. Some features were taken from metrics recommended in previous literature, while others were developed based on observed differences between the classes. A repository of all 71 evaluated features is provided in this chapter. Out of these features, 9 were selected during the feature selection process in Chapter 6. These are BillauerPeaks, ZeroCrossings, medianN, medianR, stdevE, medianACC, stdevACC, ACPeakVals1, and ACPeakVals2.

Univariate analysis of each feature was conducted to assess its class discriminability. Visualization of the class distribution of each feature was done by the construction of a class-based histogram. Then, the 5-class problem was subdivided into a series of binary class problems represented by each class combination. For each binary class combination, a receiver operating characteristic (ROC) curve was constructed, plotting the false positive rate over the true positive rate obtained from the iteration of a threshold value. The area under the curve (AUC) was computed for the ROC curve, to assess the class discriminability of the feature for each class combination. AUC values for each class combinations were compiled into a matrix for each feature.

Not that this matrix is symmetrical about the diagonal, as the AUC value for each class combination and its inverse are equal. For example, the AUC value for class 5 versus class 2 is equal to the AUC value obtained from class 2 versus class 5. The construction of the ROC curve and the calculation of the AUC value were done using the Matlab *perfcurve* function.

A value of 0.5 indicates that the feature provides no class discrimination for the corresponding binary class combination, whereas a value of 0 would indicate perfect class discrimination. As greater class discrimination is desired, a value of closer to zero is desired for each class combination.

5.1 Central Tendency and Variability Features

5.1.1 Mean of the PPG Signal (meanPPG)

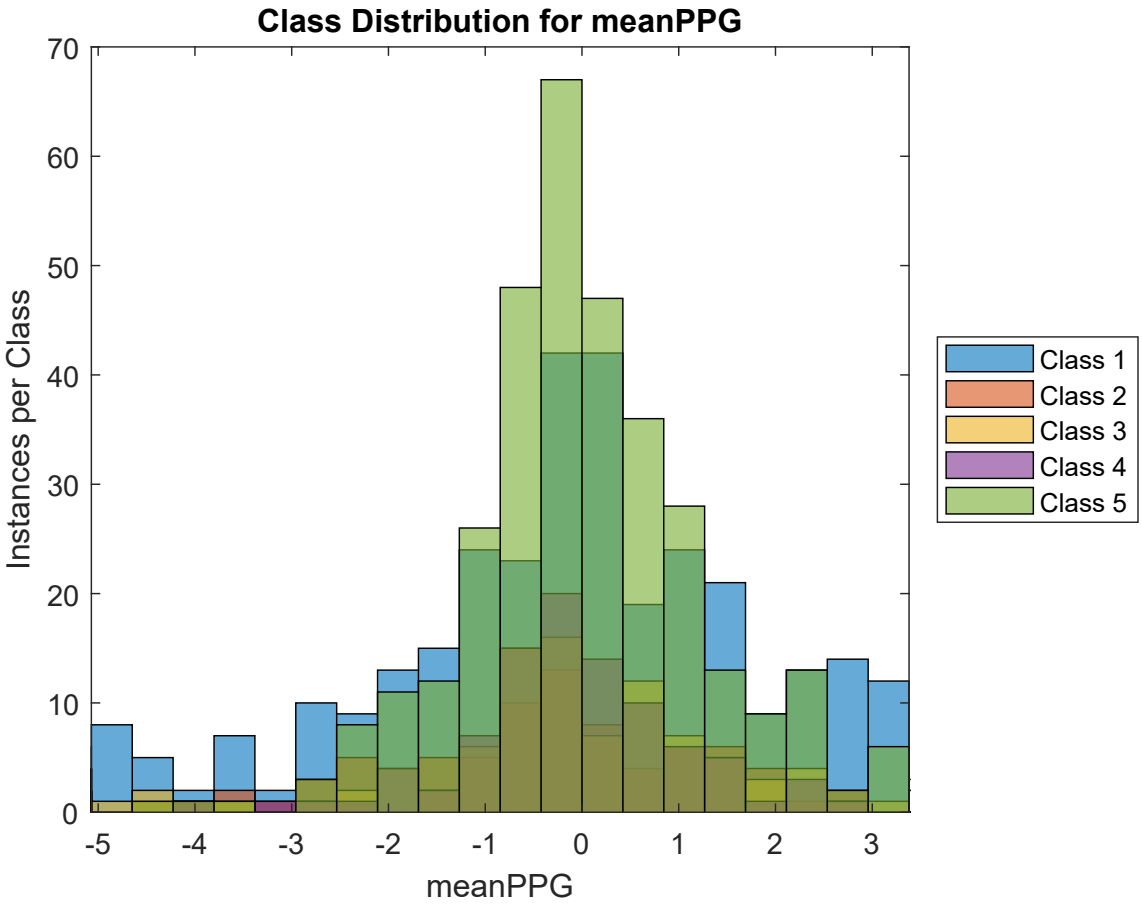

Figure 5.1: Histogram of the class distribution of the meanPPG feature.

The meanPPG feature is the mean of the PPG signal for the data segment. The feature was used to determine if there were any base differences between segments belonging to various classes. A histogram depicting the class distribution of the feature is shown in Figure 5.1. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.1.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.459	0.494	0.493	0.496
	Class 2	0.459	-	0.456	0.453	0.451
	Class 3	0.494	0.456	-	0.497	0.500
	Class 4	0.493	0.453	0.497	-	0.497
	Class 5	0.496	0.451	0.500	0.497	-

Table 5.1: Area under the curve for class versus class performance of the feature meanPPG.

5.1.2 Median of the PPG Signal (medianPPG)

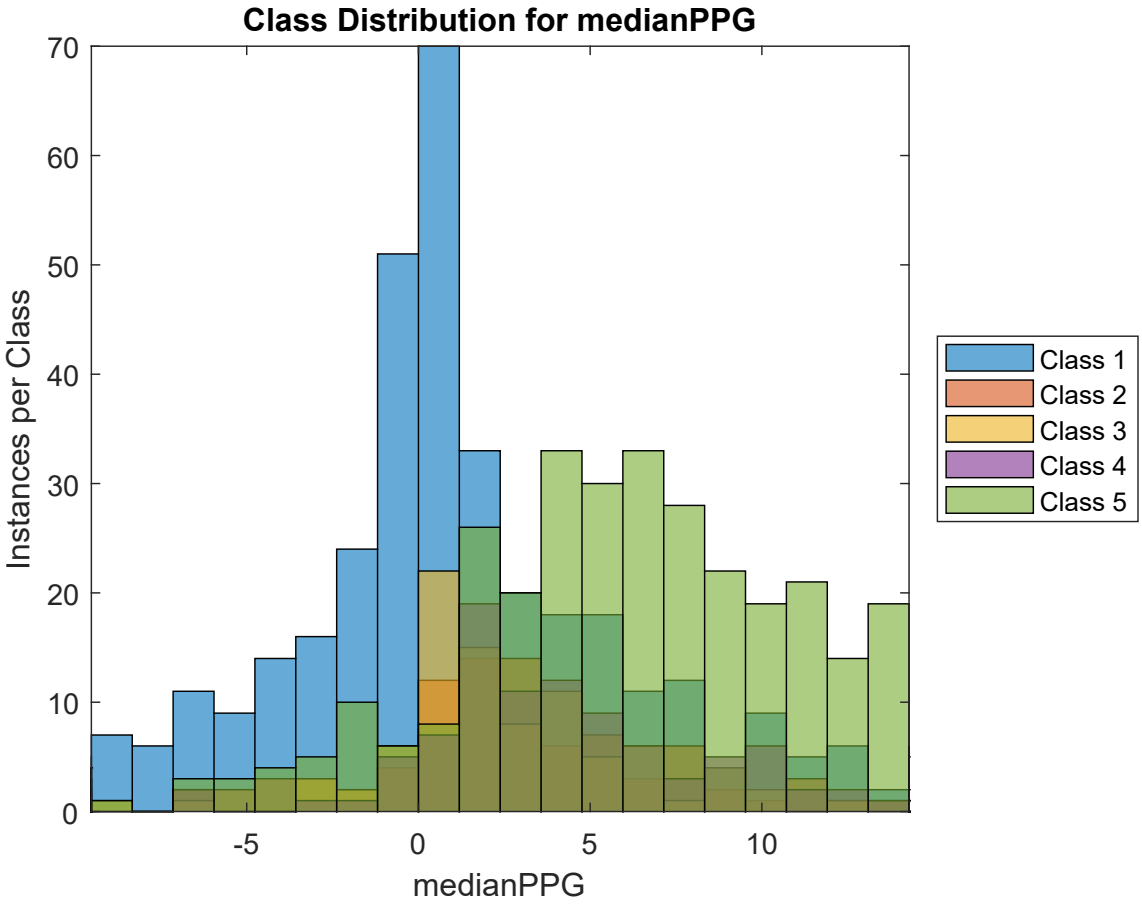

Figure 5.2: Histogram of the class distribution of the medianPPG feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.378	0.357	0.307	0.245
	Class 2	0.378	-	0.492	0.409	0.296
	Class 3	0.357	0.492	-	0.417	0.306
	Class 4	0.307	0.409	0.417	-	0.354
	Class 5	0.245	0.296	0.306	0.354	-

Table 5.2: Area under the curve for class versus class performance of the feature medianPPG.

The medianPPG feature is the median of the PPG signal for the data segment. The feature was used to determine if there were any base differences between segments belonging to various classes. A histogram depicting the class distribution of the feature is shown in Figure 5.2. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.2. The only class combinations for which the feature provides moderate discrimination is class 5 with classes 1 and 2.

5.1.3 Range of the PPG Signal (PPGRange)

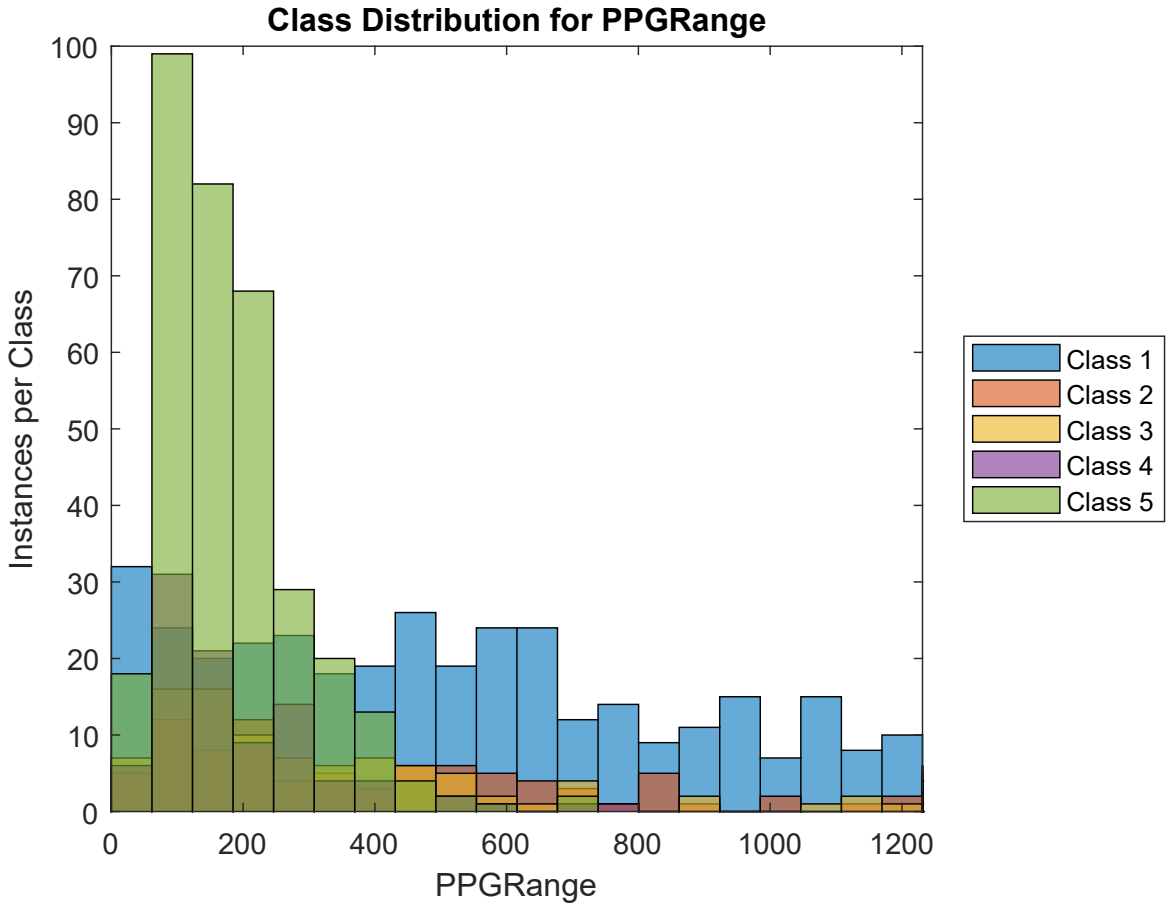

Figure 5.3: Histogram of the class distribution of the PPGRange feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.397	0.309	0.205	0.193
	Class 2	0.397	-	0.402	0.278	0.274
	Class 3	0.309	0.402	-	0.353	0.359
	Class 4	0.205	0.278	0.353	-	0.479
	Class 5	0.193	0.274	0.359	0.479	-

Table 5.3: Area under the curve for class versus class performance of the feature PPGRange.

The PPGRange feature is the total range of values in the PPG signal of the data segment. It was observed that noise corrupted PPG signals had a higher range of PPG values, hence it was postulated that this feature could aid in distinguishing the classes. The feature was computed as the difference the minimum and maximum PPG values in the data segment. A histogram depicting the class distribution of the feature is shown in Figure 5.3. According to the histogram, the feature provides good discrimination between classes 1 and 5. This is corroborated by the table of AUC values in Table 5.3. However, the feature provides poor or moderate discrimination for all other class combinations.

5.1.4 Standard Deviation of the PPG Signal (stdevPPG)

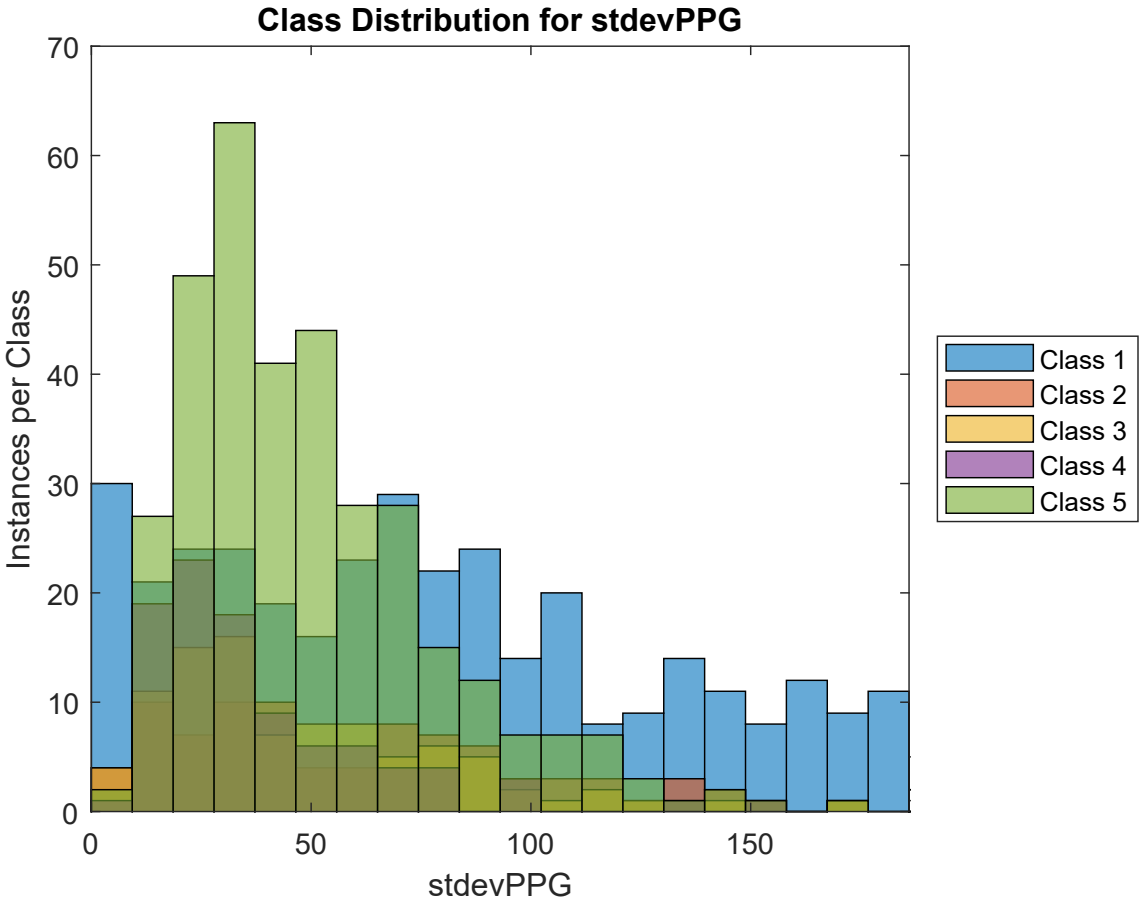

Figure 5.4: Histogram of the class distribution of the stdevPPG feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.388	0.315	0.241	0.312
	Class 2	0.388	-	0.409	0.316	0.422
	Class 3	0.315	0.409	-	0.393	0.475
	Class 4	0.241	0.316	0.393	-	0.348
	Class 5	0.312	0.422	0.475	0.348	-

Table 5.4: Area under the curve for class versus class performance of the feature stdevPPG.

The stdevPPG feature is the standard deviation of the PPG signal. It was observed that noise corrupted PPG signals had a higher range of PPG values, hence it was postulated that this feature could aid in distinguishing the classes. It was implemented in Matlab using the *std* function. A histogram depicting the class distribution of the feature is shown in Figure 5.4. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.4. The only class combinations for which the feature provides moderate discrimination is between 1 and 4.

5.1.5 Kurtosis

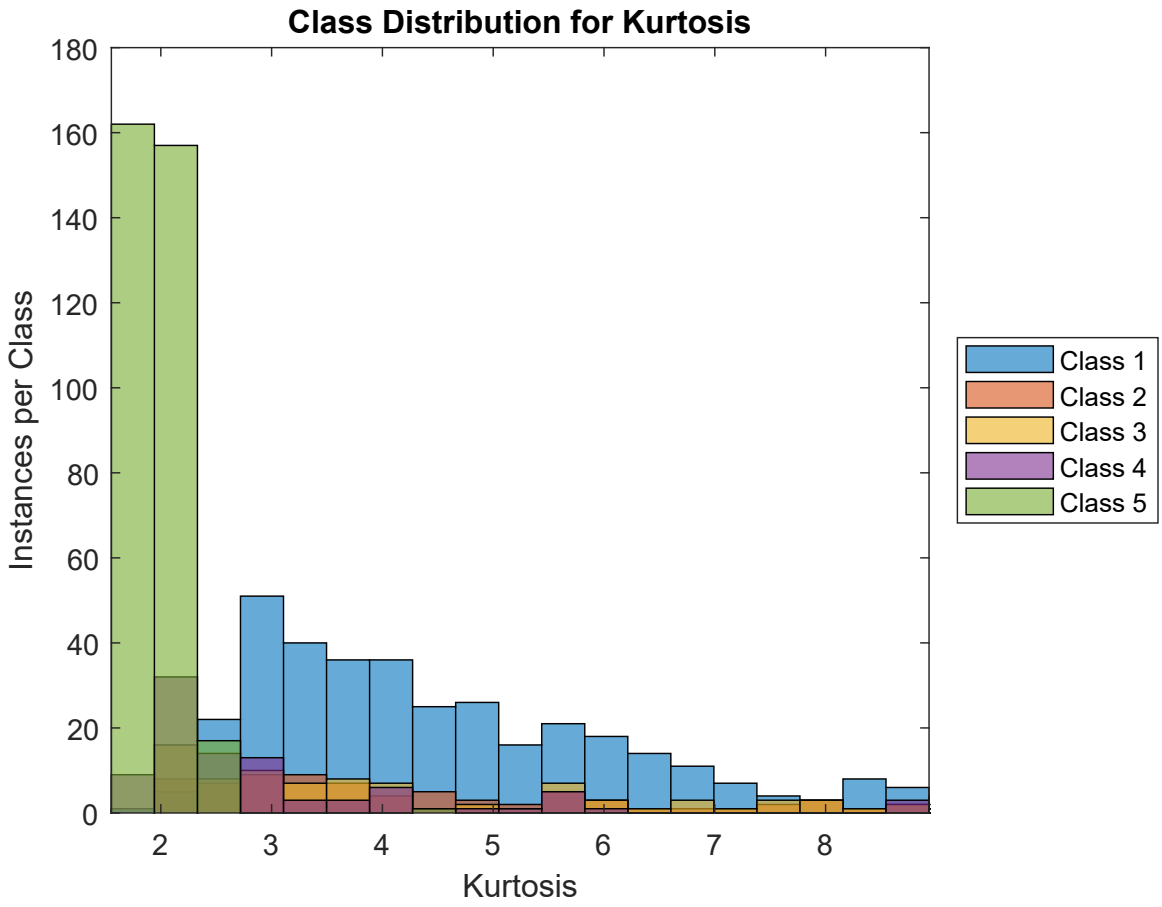

Figure 5.5: Histogram of the class distribution of the Kurtosis feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.487	0.421	0.193	0.008
	Class 2	0.487	-	0.435	0.231	0.020
	Class 3	0.421	0.435	-	0.294	0.044
	Class 4	0.193	0.231	0.294	-	0.138
	Class 5	0.008	0.020	0.044	0.138	-

Table 5.5: Area under the curve for class versus class performance of the feature Kurtosis.

The Kurtosis feature is the kurtosis of the PPG signal, which represents the distribution of the data around the mean. It is believed that this will vary according to signal quality. It was implemented in Matlab using the *kurtosis* function. This feature was inspired by the previous work of Wander and Morris [21]. A histogram depicting the class distribution of the feature is shown in Figure 5.5. According to the histogram, the feature provides good discrimination between classes 1 and 5. The table of AUC values in Table 5.5 further indicates that the feature provides good discrimination of class 5 from all other classes. Thus, this feature could be used to separate class 5 segments from the dataset.

5.1.6 Skewness

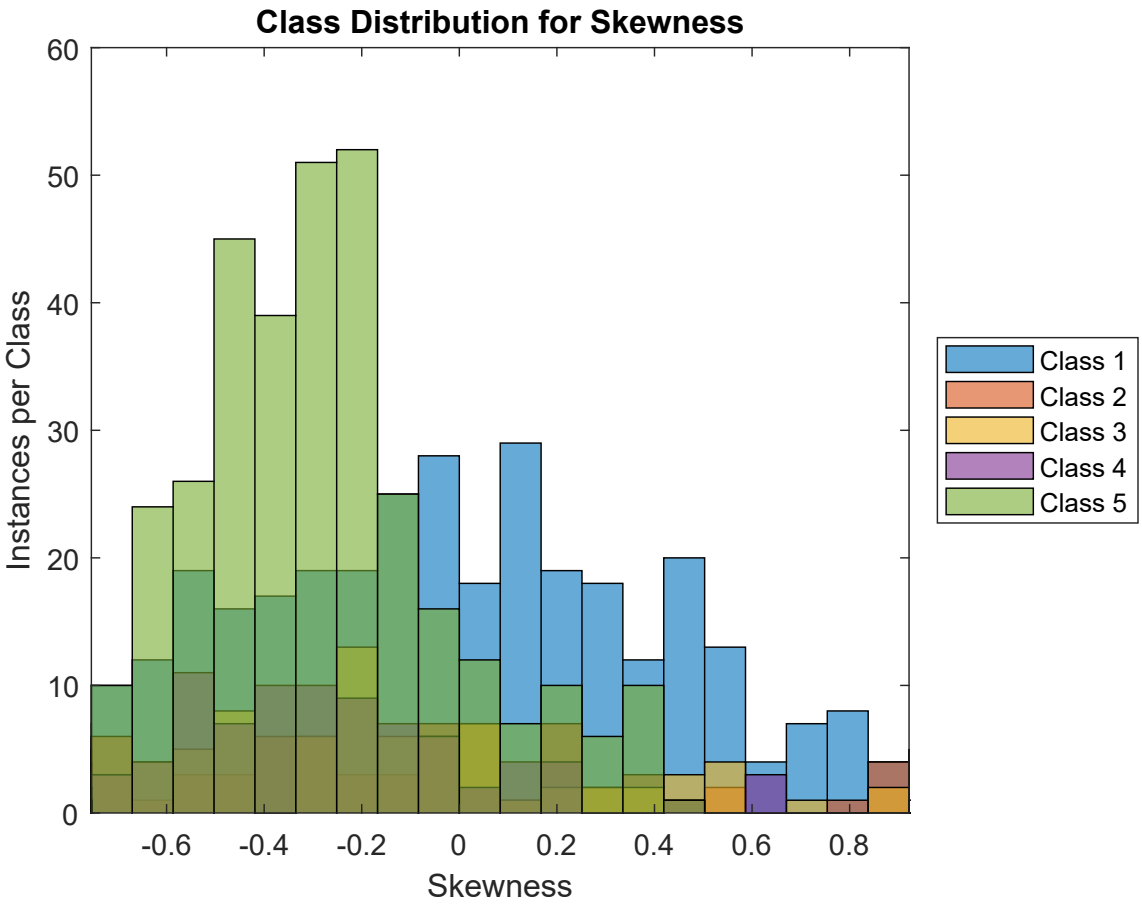

Figure 5.6: Histogram of the class distribution of the Skewness feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.450	0.410	0.391	0.374
	Class 2	0.450	-	0.470	0.456	0.447
	Class 3	0.410	0.470	-	0.483	0.473
	Class 4	0.391	0.456	0.483	-	0.492
	Class 5	0.374	0.447	0.473	0.492	-

Table 5.6: Area under the curve for class versus class performance of the feature Skewness.

The Skewness feature is the skewness of the PPG signal in the data segment. It was implemented in Matlab using the *skewness* function. The skewness is a statistic measure of the symmetry of a distribution. Data segments belonging to lower quality classes were expected to asymmetrical distributions. This feature was inspired by the previous work of Wander and Morris [21]. A histogram depicting the class distribution of the feature is shown in Figure 5.6. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.6.

5.1.7 Perfusion

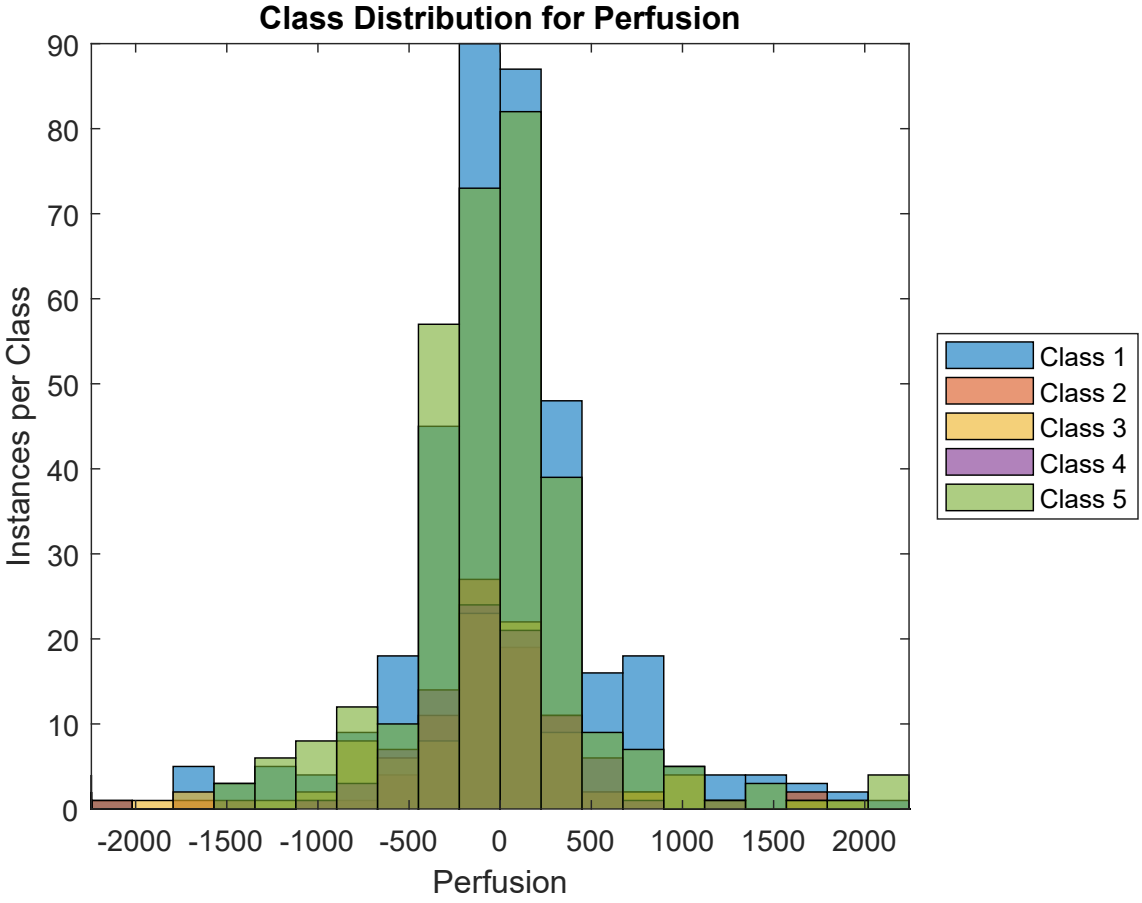

Figure 5.7: Histogram of the class distribution of the Perfusion feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.443	0.460	0.474	0.468
	Class 2	0.443	-	0.487	0.471	0.476
	Class 3	0.460	0.487	-	0.478	0.488
	Class 4	0.474	0.471	0.478	-	0.492
	Class 5	0.468	0.476	0.488	0.492	-

Table 5.7: Area under the curve for class versus class performance of the feature Perfusion.

The Perfusion feature is the skewness of the PPG signal in the data segment. It was implemented in Matlab using Equation 5.1. The perfusion is a ratio of pulsatile blood flow (AC component) to the non-pulsatile component (DC component). This feature was inspired by the previous work of Elgendi [28]. A histogram depicting the class distribution of the feature is shown in Figure 5.7. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.7.

$$Perfusion = \frac{(y_{max} - y_{min})}{\hat{x}} \times 100 \tag{5.1}$$

Where y is the PPG signal band-pass filtered in the 0.25-4Hz range, and $\hat{x}$ is the mean of the unfiltered PPG signal.

5.1.8 Noise Ratio

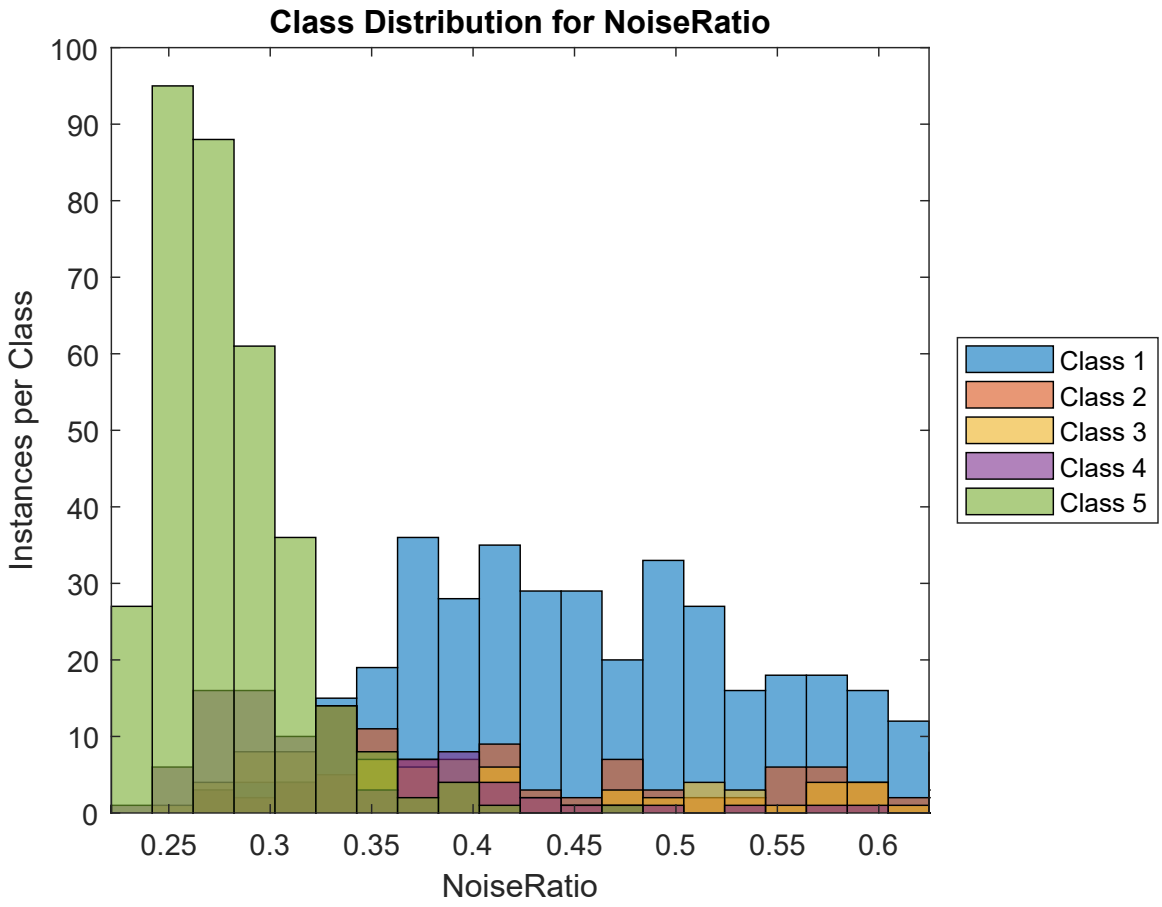

Figure 5.8: Histogram of the class distribution of the NoiseRatio feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.441	0.336	0.120	0.015
	Class 2	0.441	-	0.393	0.178	0.036
	Class 3	0.336	0.393	-	0.279	0.078
	Class 4	0.120	0.178	0.279	-	0.221
	Class 5	0.015	0.036	0.078	0.221	-

Table 5.8: Area under the curve for class versus class performance of the feature NoiseRatio.

The NoiseRatio feature is the implementation of signal-to-noise ratio (SNR) by Elgendi [28], shown in Equation 5.2. In this implementation, noise is described as the variance of the filtered PPG, while signal is defined as the variance of the absolute value of the filtered signal. The signal is band-pass filtered in the 0.25-4Hz range.

$$NoiseRatio = \frac{\sigma_y^2}{\sigma_{|y|}^2} \tag{5.2}$$

Where y is the PPG signal band-pass filtered in the 0.25-4Hz range.

A histogram depicting the class distribution of the feature is shown in Figure 5.8. According to the histogram, the feature provides good discrimination between class 1 and 5. Additionally, the table of AUC values in Table 5.8 indicates that the feature offers good discrimination of class 4 from classes 1 and 2, as well as class 5 from classes 1, 2, and 3.

5.1.9 Zero-Crossing Rate (ZeroCrossings)

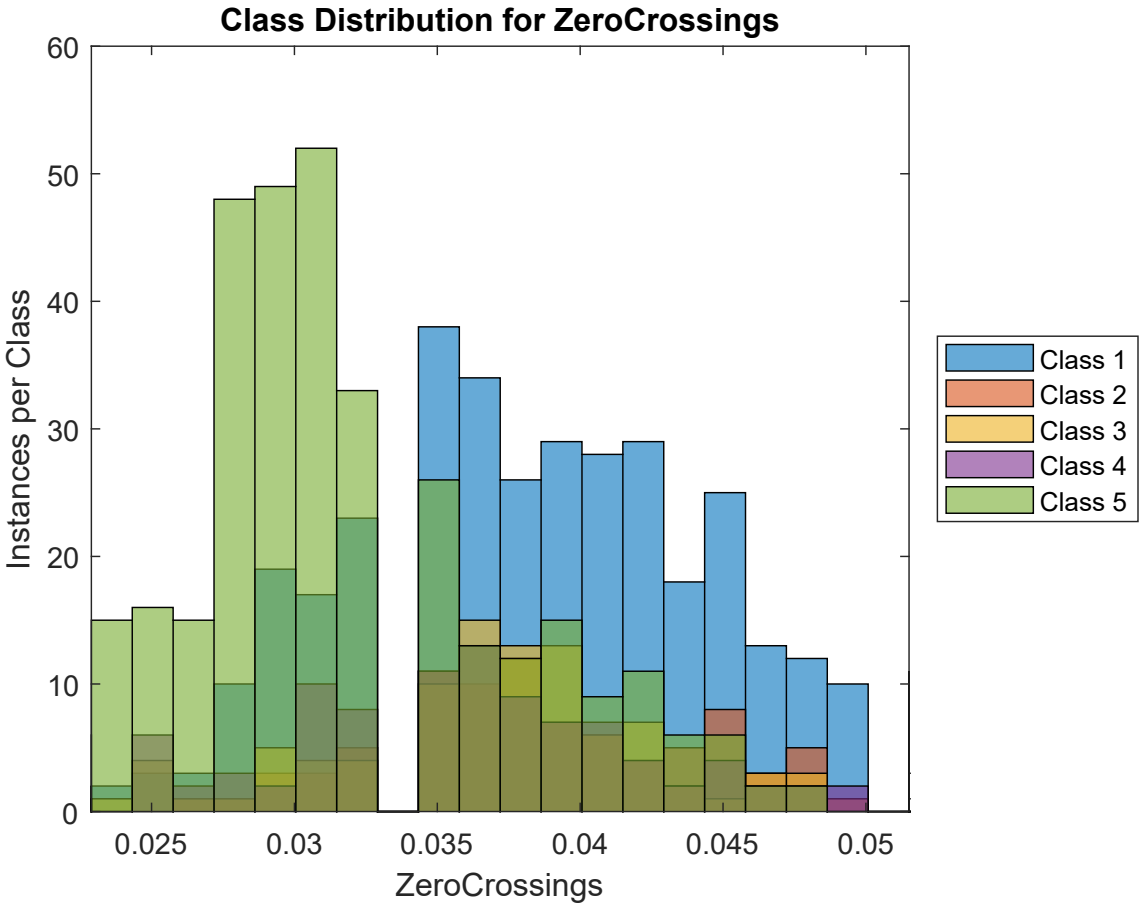

Figure 5.9: Histogram of the class distribution of the ZeroCrossings feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.467	0.393	0.346	0.194
	Class 2	0.467	-	0.419	0.361	0.189
	Class 3	0.393	0.419	-	0.436	0.249
	Class 4	0.346	0.361	0.436	-	0.302
	Class 5	0.194	0.189	0.249	0.302	-

Table 5.9: Area under the curve for class versus class performance of the feature ZeroCrossings.

The ZeroCrossings feature is the number of times the PPG signal crosses the x-axis ($y = 0$), divided by the length of the data segment. It was expected that data segments belonging to lower classes would have a higher zero-crossing rate. This feature was inspired by the previous work of Elgendi [28]. A histogram depicting the class distribution of the feature is shown in Figure 5.9. According to the histogram, the feature provides some discrimination between class 1 and 5. The table of AUC values in Table 5.9 shows that the feature provides good discrimination of class 5 from classes 1 and 2.

5.2 Miscellaneous Signal Features

5.2.1 Entropy

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.407	0.325	0.260	0.318
	Class 2	0.407	-	0.395	0.313	0.404
	Class 3	0.325	0.395	-	0.411	0.478
	Class 4	0.260	0.313	0.411	-	0.369
	Class 5	0.318	0.404	0.478	0.369	-

Table 5.10: Area under the curve for class versus class performance of the feature Entropy.

Entropy is a measure of the uncertainty present in the PPG signal, using the difference between the probability density of the signal from a uniform distribution [28]. Lower quality classes were expected to result in higher entropy. The feature was inspired by the work of Elgendi [28], and computed using Equation 5.3.

$$Entropy = -\sum_{n=1}^{N} \Big((x[n]^2) \Big(log_e(x[n]^2) \Big) \Big) \tag{5.3}$$

A histogram depicting the class distribution of the feature is shown in Figure 5.10. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.10.

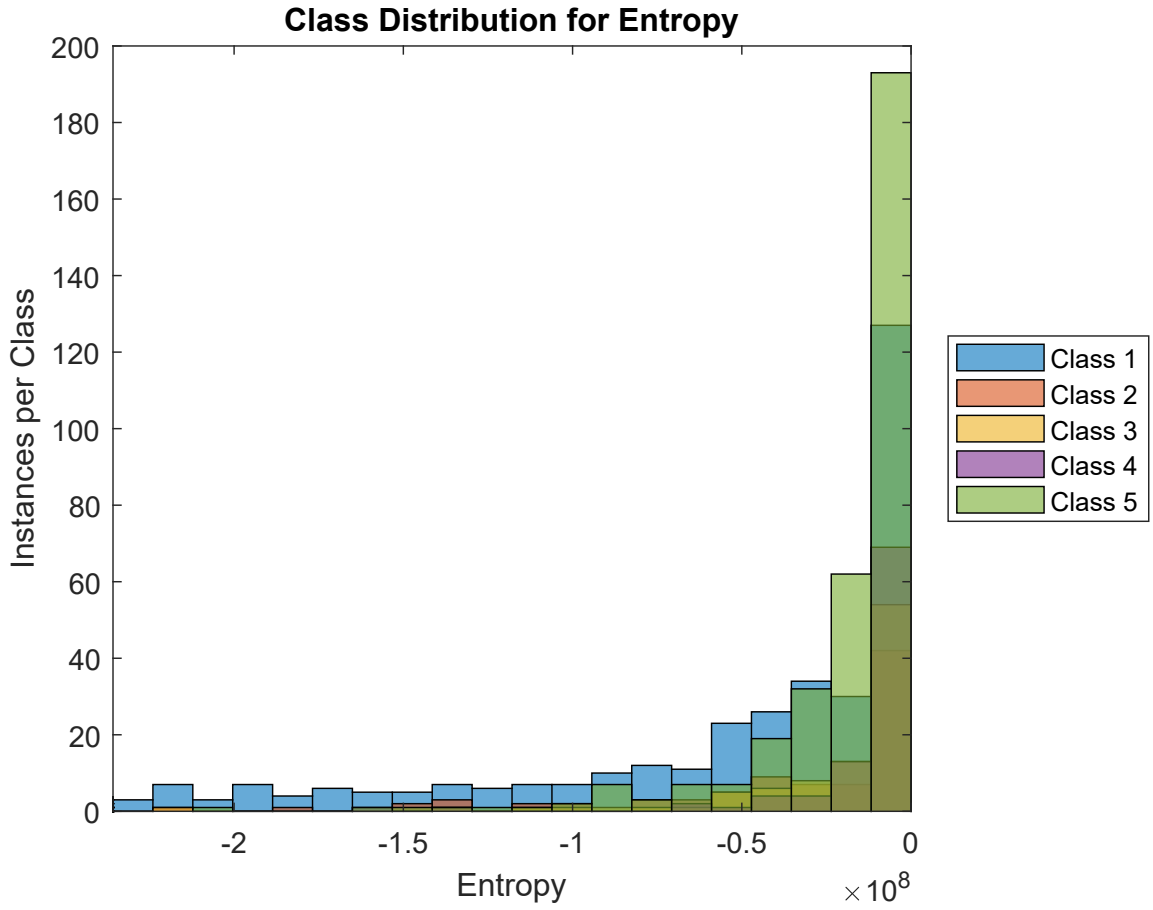

Figure 5.10: Histogram of the class distribution of the Entropy feature.

5.2.2 Number of Peaks Identified by Billauer's Algorithm (BillauerPeaks)

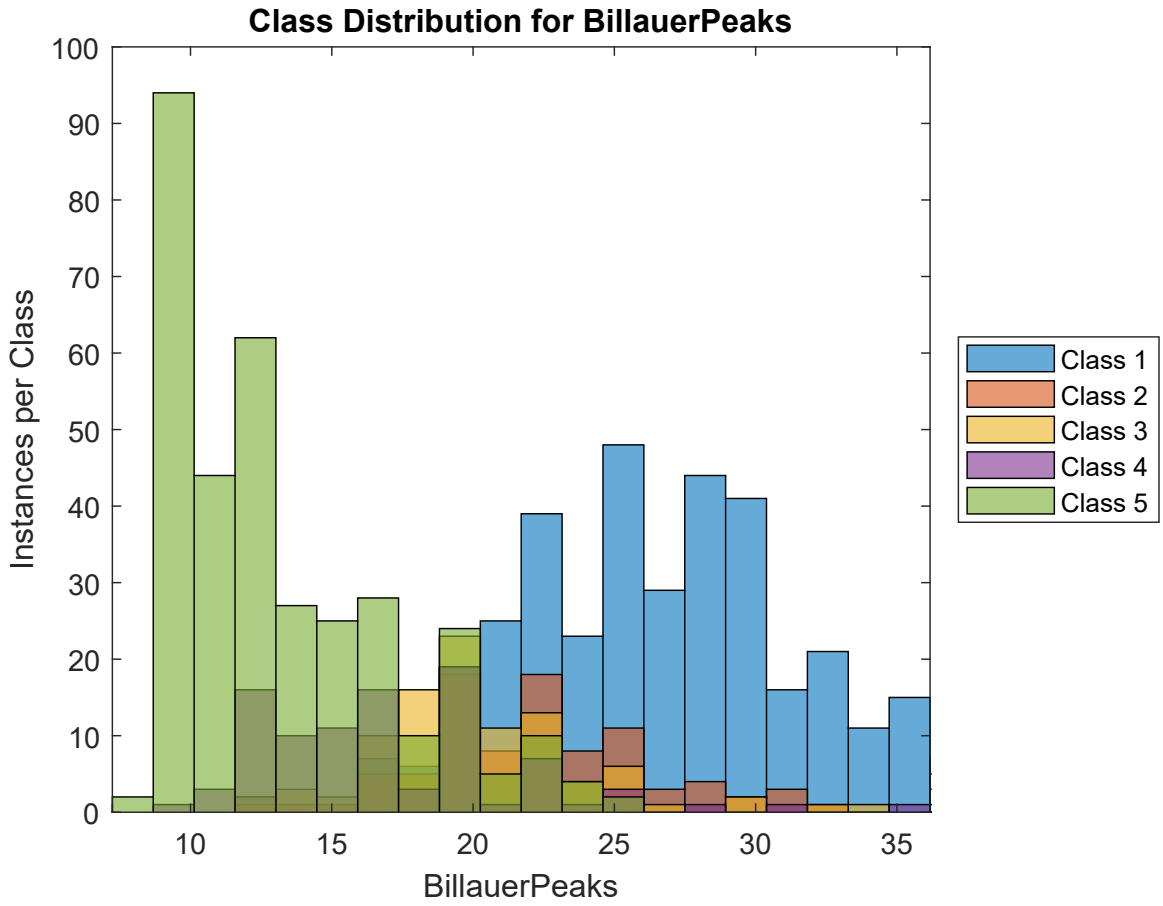

Figure 5.11: Histogram of the class distribution of the BillauerPeaks feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.237	0.136	0.079	0.021
	Class 2	0.237	-	0.334	0.179	0.063
	Class 3	0.136	0.334	-	0.273	0.108
	Class 4	0.079	0.179	0.273	-	0.242
	Class 5	0.021	0.063	0.108	0.242	-

Table 5.11: Area under the curve for class versus class performance of the feature BillauerPeaks.

The BillauerPeaks feature is the number of peaks in the PPG signal identified using Billauer's algorithm for peak detection. Rather than using a derivative-based approach, the algorithm identifies peaks based on adjacent values being lower than a specific threshold [29]. This was implement in Matlab using the *peakdet* function from `http://www.billauer.co.il/peakdet.html`. As seen in Figures 5.12 and 5.13, a greater number of peaks were expected to be found in data belonging to lower quality classes.

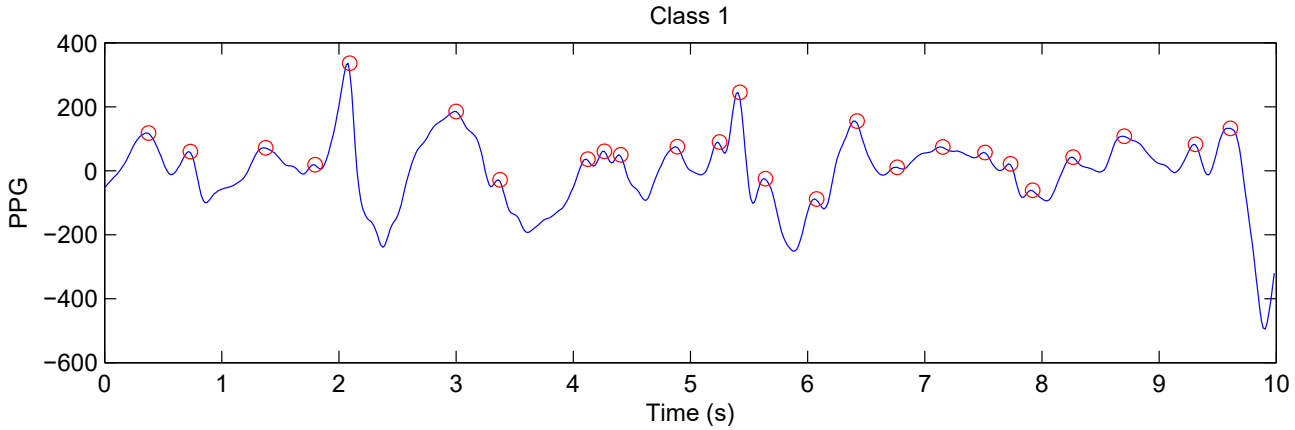

Figure 5.12: Example of peaks found using Billauer's algorithm in a class 1 PPG segment.

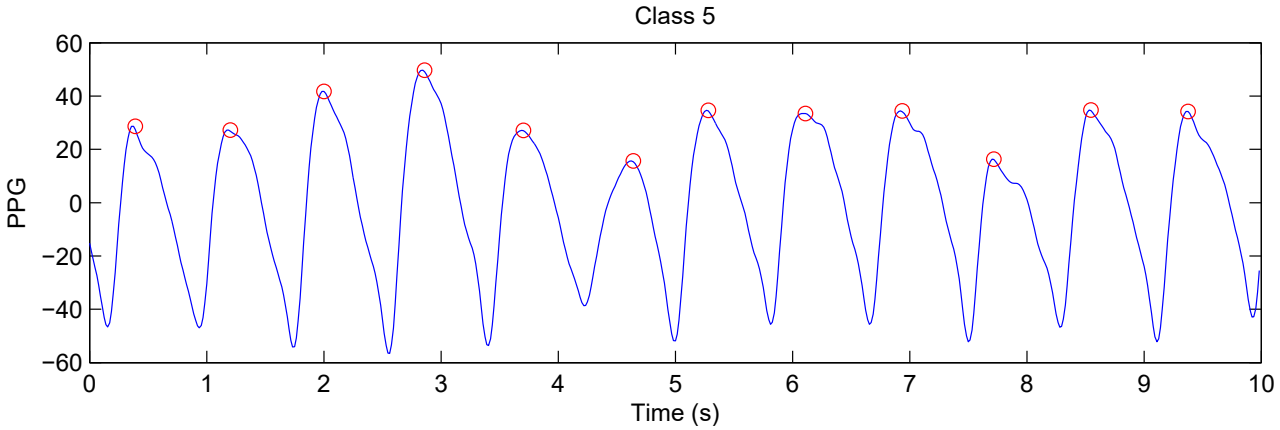

Figure 5.13: Example of peaks found using Billauer's algorithm in a class 5 PPG segment.

A histogram depicting the class distribution of the feature is shown in Figure 5.11. According to the histogram, the feature provides good discrimination between class 1 and 5. The table of AUC values in Table 5.11 shows that the feature provides good or moderate discrimination for all class combinations, with the exception of the class 3 and 4 combination. Thus, BillauerPeaks is a strong feature.

5.2.3 Zero-Crossings of the Instantaneous Frequency

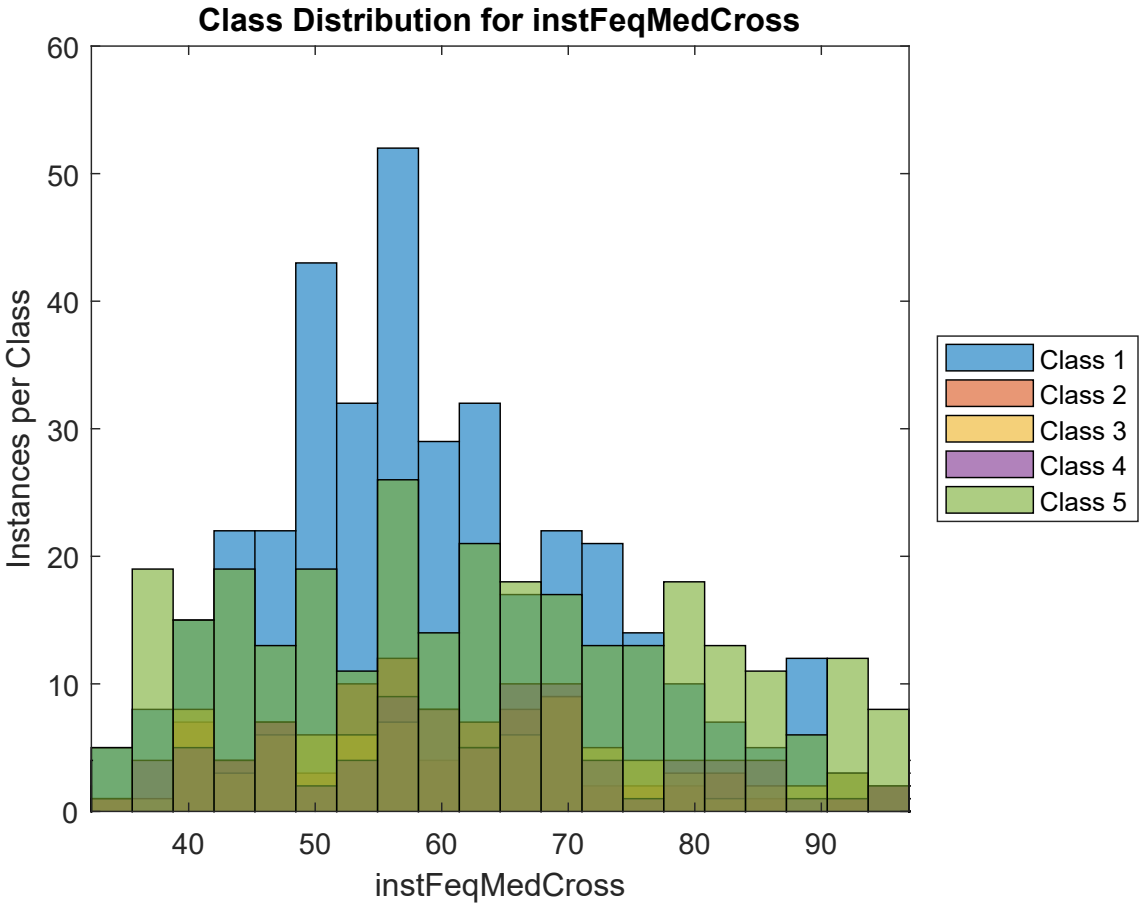

Figure 5.14: Histogram of the class distribution of the instFeqMedCross feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.498	0.463	0.442	0.434
	Class 2	0.498	-	0.462	0.440	0.433
	Class 3	0.463	0.462	-	0.484	0.466
	Class 4	0.442	0.440	0.484	-	0.478
	Class 5	0.434	0.433	0.466	0.478	-

Table 5.12: Area under the curve for class versus class performance of the feature instFeqMedCross.

The Hilbert Transform was taken to obtain the instantaneous frequency of the PPG signal. It was postulated that lower quality classes would produce noisier instantaneous frequency plots, thus the number of median crossings were used as a feature. The use of this feature was inspired by preliminary work conducted by the research group [2].

A histogram depicting the class distribution of the feature is shown in Figure 5.14. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.12.

5.3 Poincaré Plot Features

Poincaré plots are used to visualize reoccurred and predictability in a signal. The plots are constructed by plotting $x(n)$ over $x(n+1)$, where x represents the PPG signal. Predictable, periodic data results in a scatter plot in the shape of an ellipse. Class 5 segments were postulated to have Poincaré plot data points shaped closer to an ellipse, compared to segments belong to lower quality classes. An ellipse is fitted to the Poincaré plot using the *EllipseDirectFit* function in Matlab [30], which is based on the direct least squares fitting method [31]. From the fitted ellipse, major and minor axis lengths are determined and used as features. Given an ellipse represented by Equation 5.4, the semi-axis lengths are represented by Equations 5.5 and 5.6 [32]. The longer length is used as the semi-major axis length and the shorter length is used as the semi-minor axis length.

$$ax^2 + 2bxy + cy^2 + 2dx + 2fy + g = 0 \tag{5.4}$$

$$L_1 = \sqrt{\frac{2(af^2 + cd^2 + gb^2 - 2bdf - acg)}{(b^2 - ac)\left[\sqrt{(a-c)^2 + 4b^2} - (a+c)\right]}} \tag{5.5}$$

$$L_2 = \sqrt{\frac{2(af^2 + cd^2 + gb^2 - 2bdf - acg)}{(b^2 - ac)\left[-\sqrt{(a-c)^2 + 4b^2} - (a+c)\right]}} \tag{5.6}$$

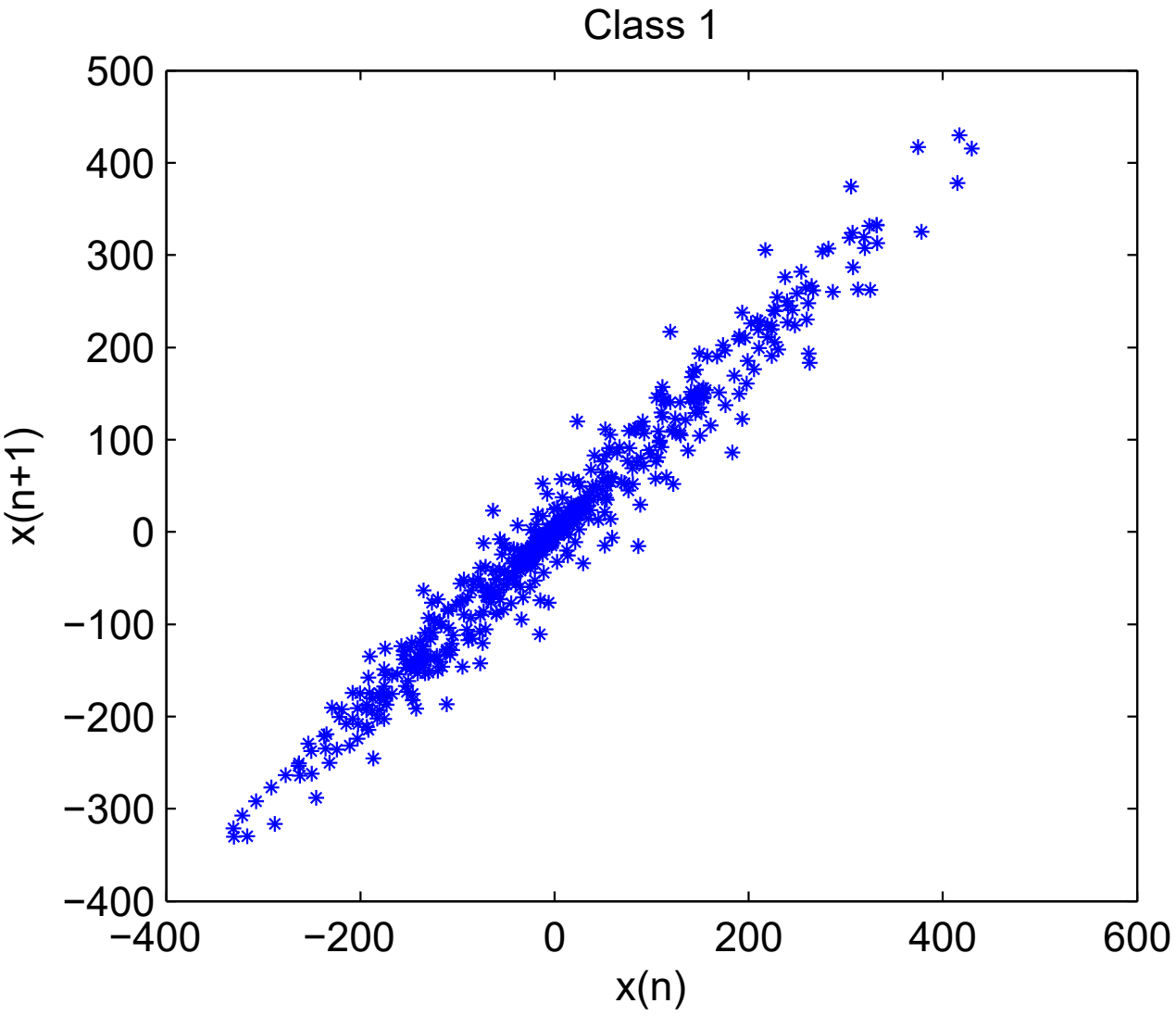

Figure 5.15: Example of Poincaré plot for class 1 data segment. x represents the PPG signal.

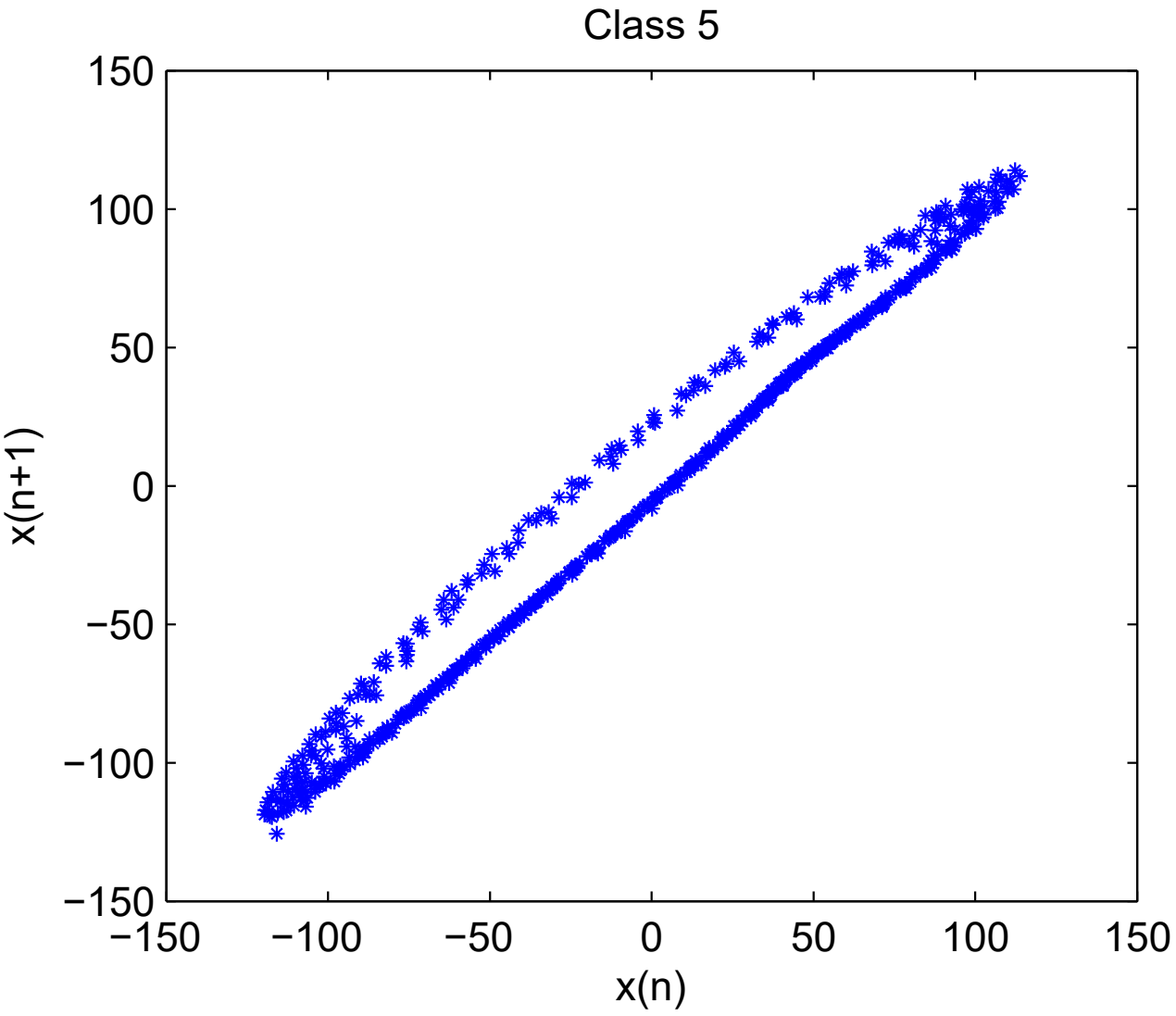

Figure 5.16: Example of Poincaré plot for class 5 data segment. x represents the PPG signal.

5.3.1 Length of the Major Axis (sMajor)

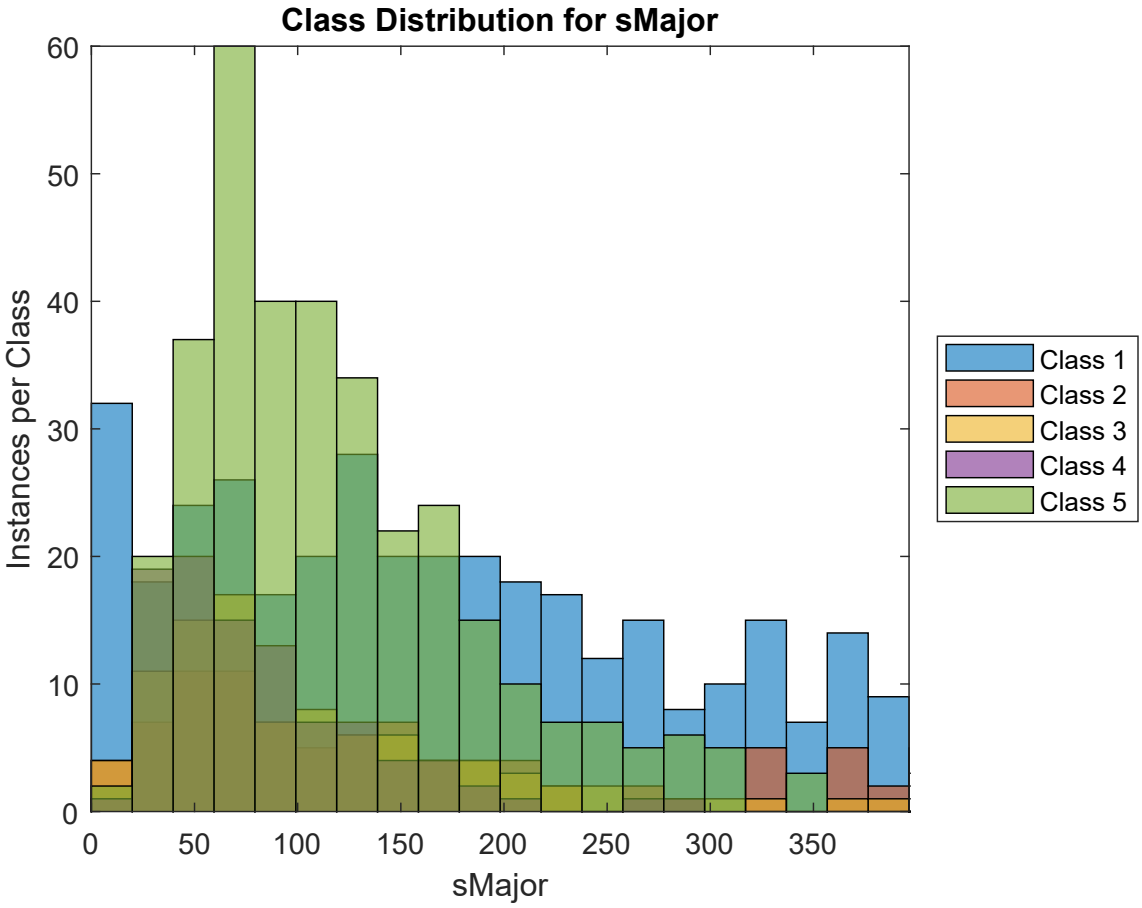

Figure 5.17: Histogram of the class distribution of the sMajor feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.405	0.323	0.251	0.348
	Class 2	0.405	-	0.413	0.329	0.460
	Class 3	0.323	0.413	-	0.410	0.432
	Class 4	0.251	0.329	0.410	-	0.327
	Class 5	0.348	0.460	0.432	0.327	-

Table 5.13: Area under the curve for class versus class performance of the feature sMajor.

The sMajor feature is length of the semi-major axis obtained from the fitting of an ellipse to the Poincaré plot. A histogram depicting the class distribution of the feature is shown in Figure 5.17. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.13. The only combination for which there is moderate discrimination is between classes 1 and 4.

5.3.2 Length of the Minor Axis (sMinor)

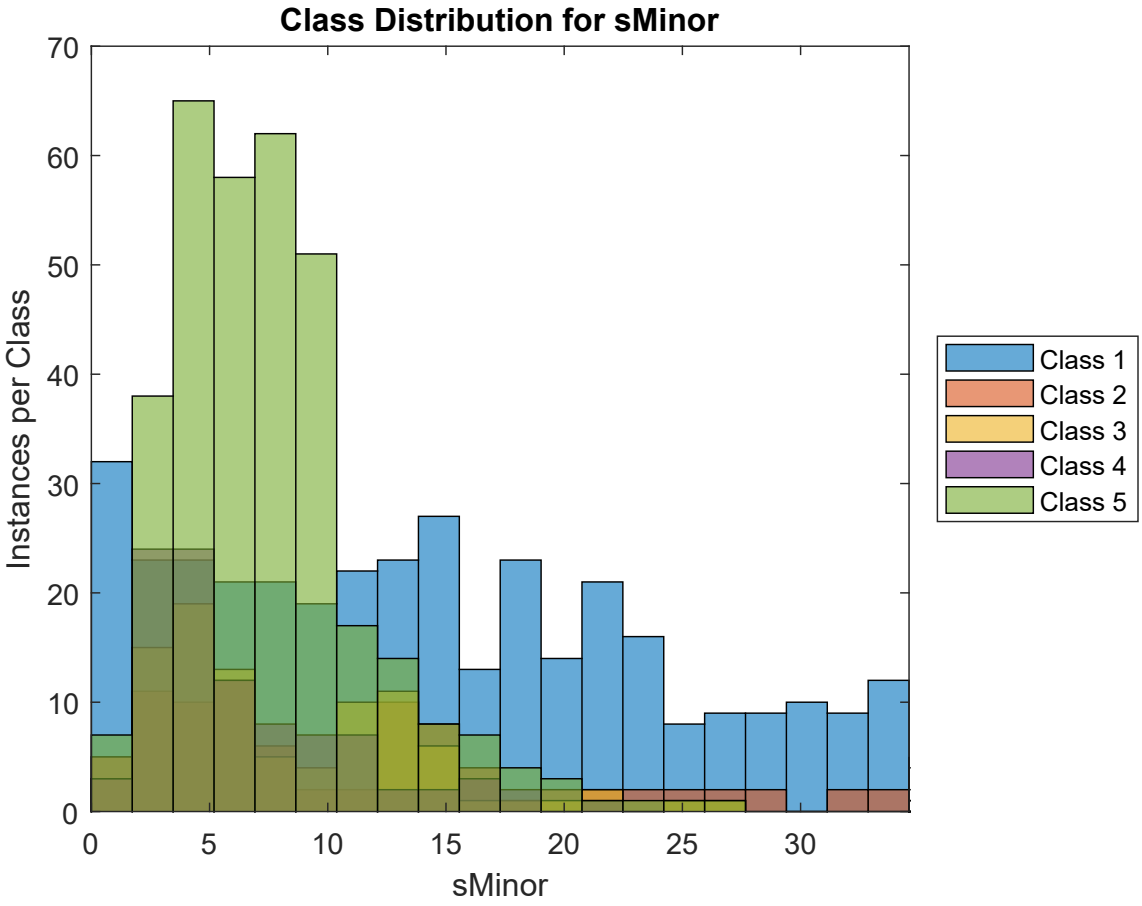

Figure 5.18: Histogram of the class distribution of the sMinor feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.393	0.302	0.232	0.267
	Class 2	0.393	-	0.402	0.319	0.389
	Class 3	0.302	0.402	-	0.393	0.480
	Class 4	0.232	0.319	0.393	-	0.388
	Class 5	0.267	0.389	0.480	0.388	-

Table 5.14: Area under the curve for class versus class performance of the feature sMinor.

The sMinor feature is length of the semi-major axis obtained from the fitting of an ellipse to the Poincaré plot. A histogram depicting the class distribution of the feature is shown in Figure 5.18. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.14. The only combinations for which there is moderate discrimination is between classes 1 with classes 4 and 5.

5.4 Average Pulse Features

An average pulse is computed for each data segment by identifying all the pulses in the segment, and taking their mean. The pulses were identified using Billauer's algorithm for peak detection, implemented with the custom Matlab function *peakdet* [29]. Each detected peak is assumed to represent the peak of a PPG pulse, as seen in Figure 5.13. The individual PPG pulses are segmented by taking a fixed window of 61 data points (approximately 0.95 seconds), with 18 data points before the peak, and 42 data points following the peak. This window size was manually optimized on a small segment of visually identified class 5 PPG data. The uneven window length on either side of the peak reflects the morphology of the ideal PPG pulse, which is characterized by a quick rise to the peak, followed by a slow fall. The identified pulses within a segment were averaged to obtain the average pulse for each data segment. In class 5 data segments, the average pulse ought to resemble an ideal PPG pulse. However, segments belonging to lower quality classes were expected to have peaks identified in signal noise, as seen in Figure 5.12, hence the average pulse from those classes would have a different morphology. The following statistics are computed on the average pulse computer for each segment.

5.4.1 Entropy of Average Pulse (EPulseAvg)

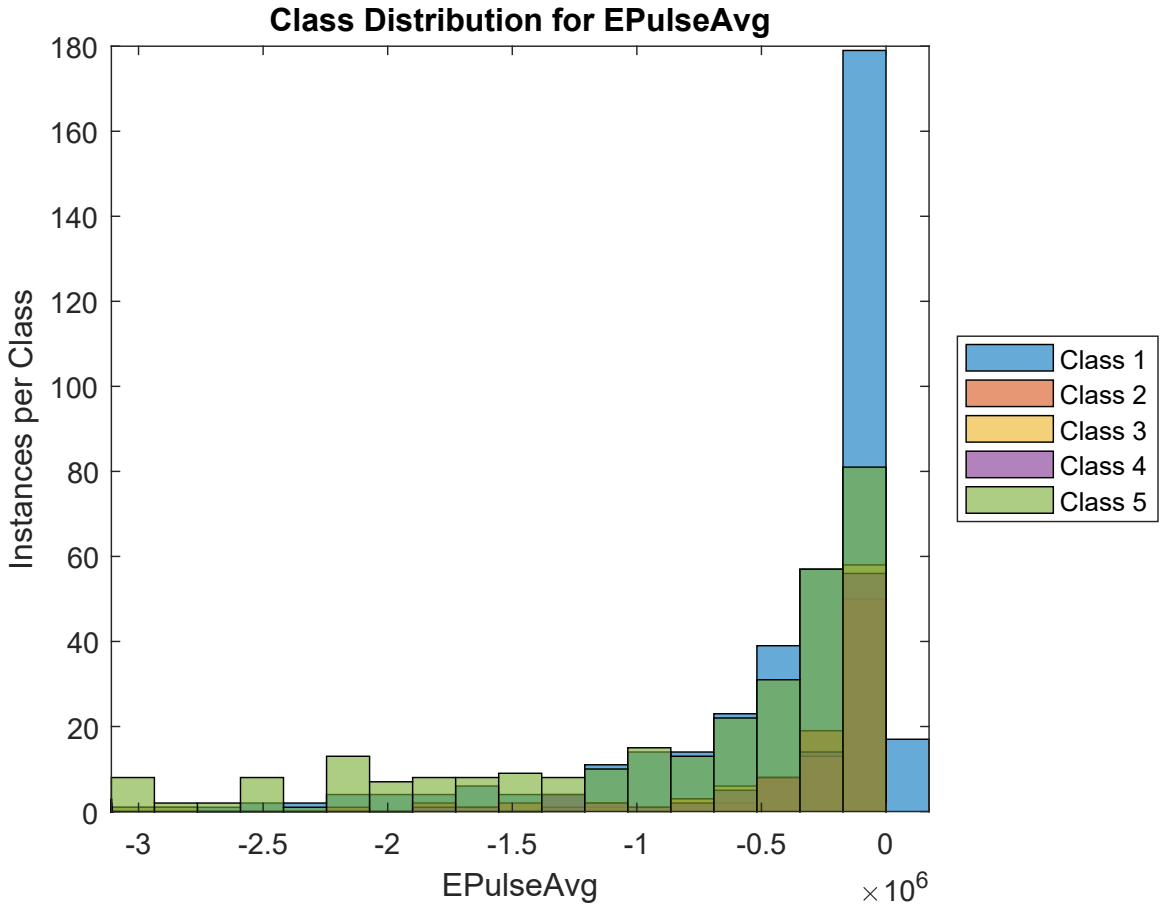

Figure 5.19: Histogram of the class distribution of the EPulseAvg feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.461	0.443	0.468	0.304
	Class 2	0.461	-	0.480	0.494	0.249
	Class 3	0.443	0.480	-	0.470	0.235
	Class 4	0.468	0.494	0.470	-	0.252
	Class 5	0.304	0.249	0.235	0.252	-

Table 5.15: Area under the curve for class versus class performance of the feature EPulseAvg.

The EPulseAvg feature is the entropy of the average pulse from the PPG segment. Entropy is a measure of the uncertainty present in the PPG signal, using the difference between the probability density of the signal from a uniform distribution [28]. It was computed using Equation 5.3.

A histogram depicting the class distribution of the feature is shown in Figure 5.19. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.15, though the feature does provide moderate discrimination of class 5 from classes 2, 3, and 4.

5.4.2 Kurtosis of Average Pulse (KPulseAvg)

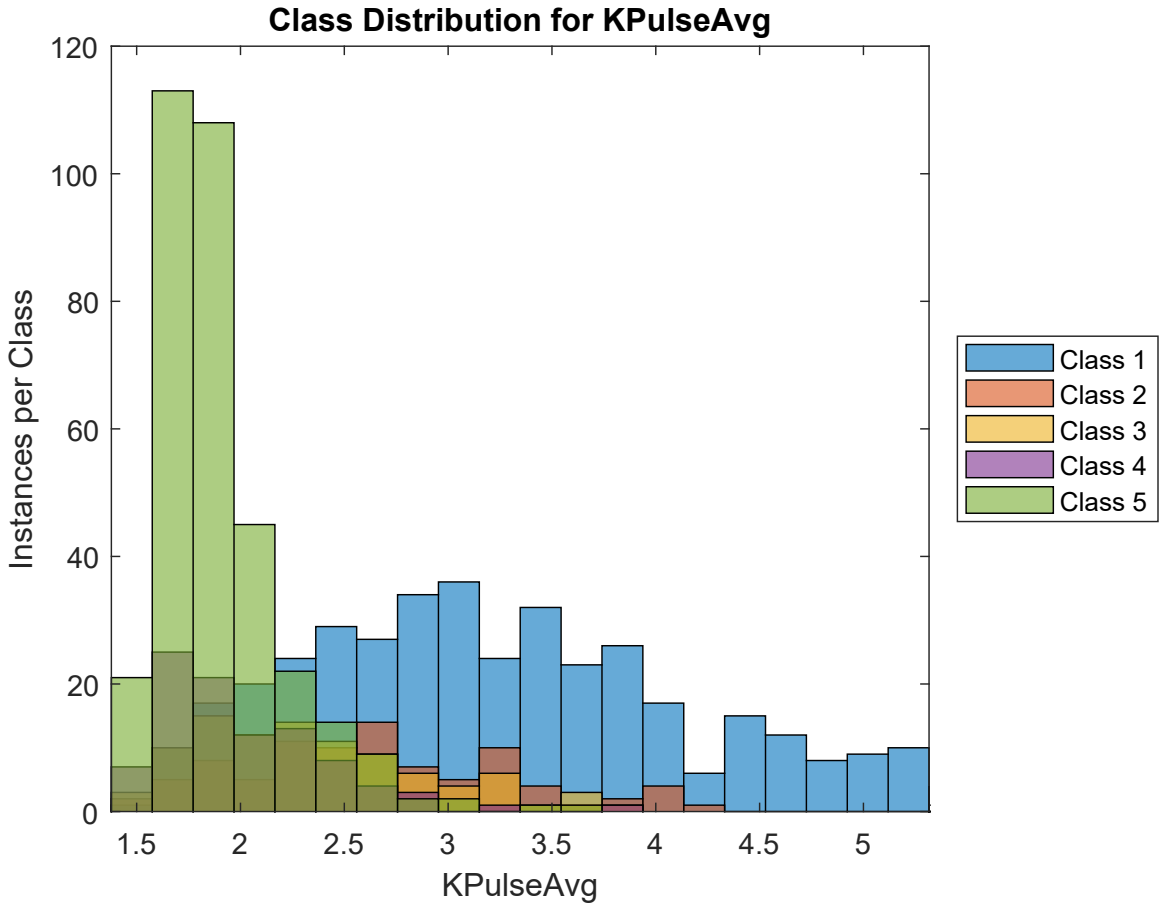

Figure 5.20: Histogram of the class distribution of the KPulseAvg feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.317	0.209	0.102	0.067
	Class 2	0.317	-	0.359	0.194	0.139
	Class 3	0.209	0.359	-	0.310	0.236
	Class 4	0.102	0.194	0.310	-	0.421
	Class 5	0.067	0.139	0.236	0.421	-

Table 5.16: Area under the curve for class versus class performance of the feature KPulseAvg.

The KPulseAvg feature is the kurtosis of the average pulse from the PPG segment. Kurtosis represents the distribution of the data around the mean. It is believed that this will vary according to signal quality. It was implemented in Matlab using the *kurtosis* function.

A histogram depicting the class distribution of the feature is shown in Figure 5.20. According to the histogram, the feature provides good discrimination discrimination between classes 1 and 5. The table of AUC values in Table 5.16 shows that the feature also provides good discrimination of classes 1 and 2 from both class 4 and class 5.

5.4.3 Noise Ratio of Average Pulse (NPulseAvg)

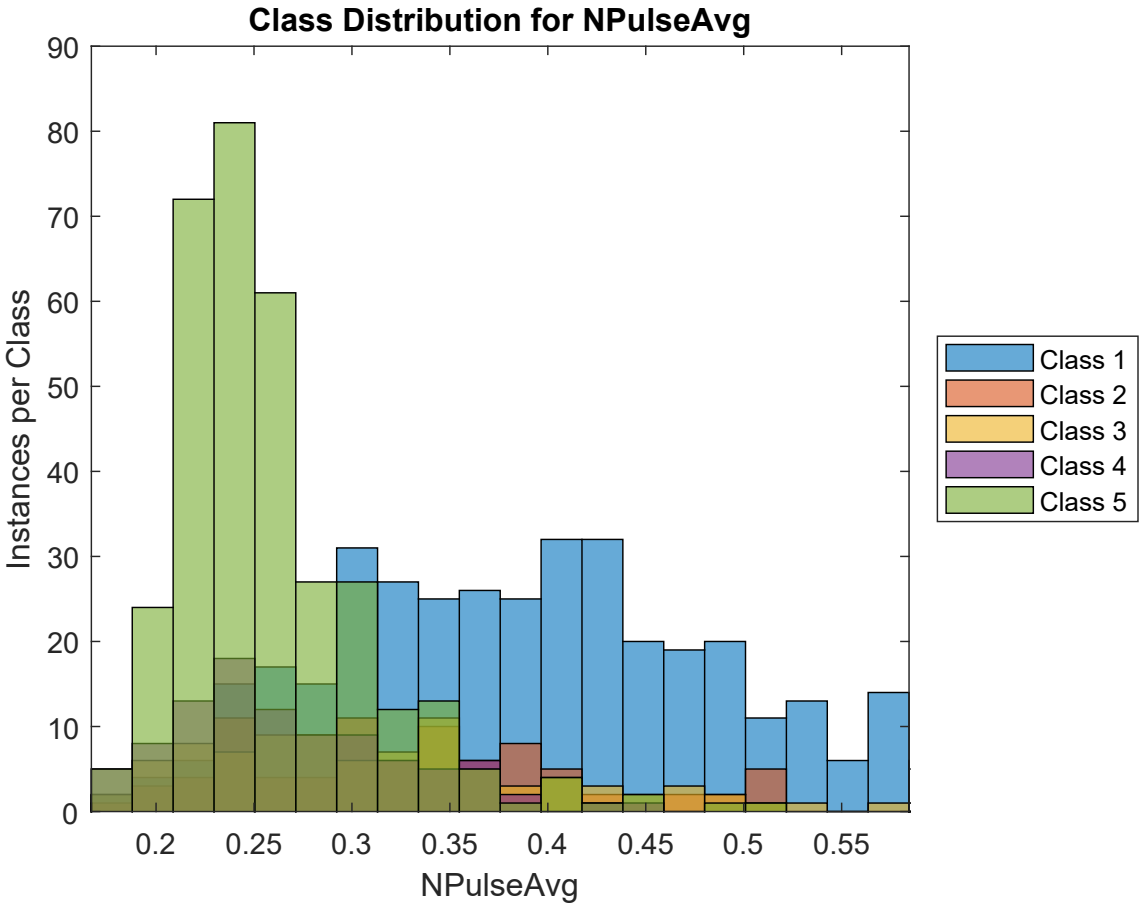

Figure 5.21: Histogram of the class distribution of the NPulseAvg feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.333	0.283	0.137	0.104
	Class 2	0.333	-	0.433	0.272	0.223
	Class 3	0.283	0.433	-	0.346	0.302
	Class 4	0.137	0.272	0.346	-	0.444
	Class 5	0.104	0.223	0.302	0.444	-

Table 5.17: Area under the curve for class versus class performance of the feature NPulseAvg.

The NPulseAvg feature is the signal-to-noise ratio of the average pulse from the PPG segment, implemented using the method by Elgendi [28] in Equation 5.2. The implementation here differs from the implementation of the NoiseRatio feature, in that the signal is not filtered. The high-pass component of the band-pass filtering was deemed unnecessary as baseline drift was not expected to be present in the short window length of the PPG pulses. The low-pass component was also deemed unnecessary as the presence of high frequencies were expected to be a good indicator of low signal quality. Hence, Equation 5.2 was implemented using the unfiltered average pulse.

A histogram depicting the class distribution of the feature is shown in Figure 5.21. According to the histogram, the feature provides good discrimination discrimination between classes 1 and 5. Additionally, the table of AUC values in Table 5.17 indicates that the feature provides good discrimination between classes 1 and 4. Other class combinations yield poor or moderate class discrimination.

5.4.4 Correlation Coefficient of Average Pulse with Template (rPulseAvg)

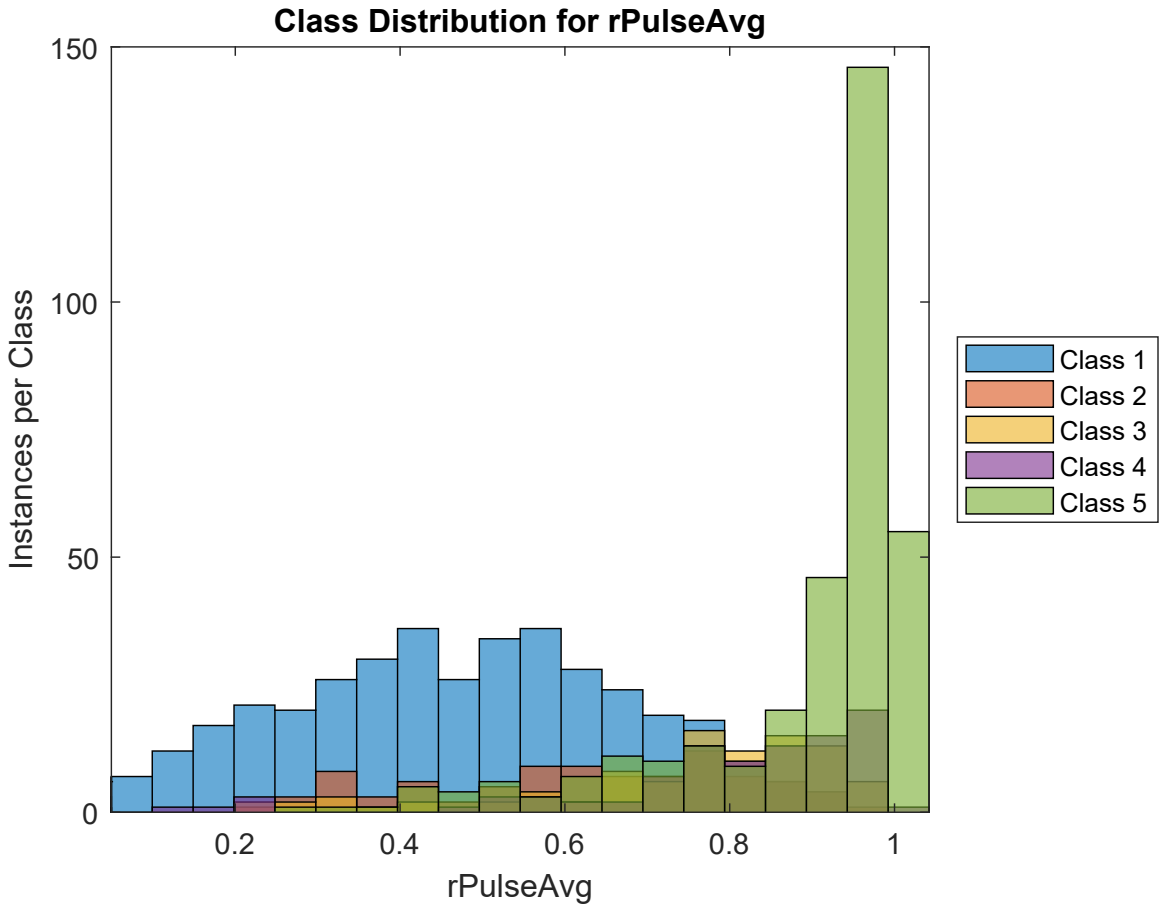

Figure 5.22: Histogram of the class distribution of the rPulseAvg feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.284	0.152	0.105	0.049
	Class 2	0.284	-	0.302	0.219	0.107
	Class 3	0.152	0.302	-	0.383	0.189
	Class 4	0.105	0.219	0.383	-	0.274
	Class 5	0.049	0.107	0.189	0.274	-

Table 5.18: Area under the curve for class versus class performance of the feature rPulseAvg.

The rPulseAvg feature is the correlation coefficient between the averaged pulse and a template of high quality PPG pulses. The template of high quality PPG pulses was created by visual identification of 10 pulses, which were then averaged. The correlation coefficient between the averaged pulse and the template was found using the *corrcoef* function in Matlab.

A histogram depicting the class distribution of the feature is shown in Figure 5.22. According to the histogram, the feature provides good discrimination discrimination between classes 1 and 5. Additionally, the table of AUC values in Table 5.18 indicates that the feature provides good discrimination of class 1 from classes 3, 4, and 5, and of class 5 from classes 1, 2, and 3. Other class combinations yield poor or moderate discrimination.

5.4.5 Relative Power of Average Pulse (RPulseAvg)

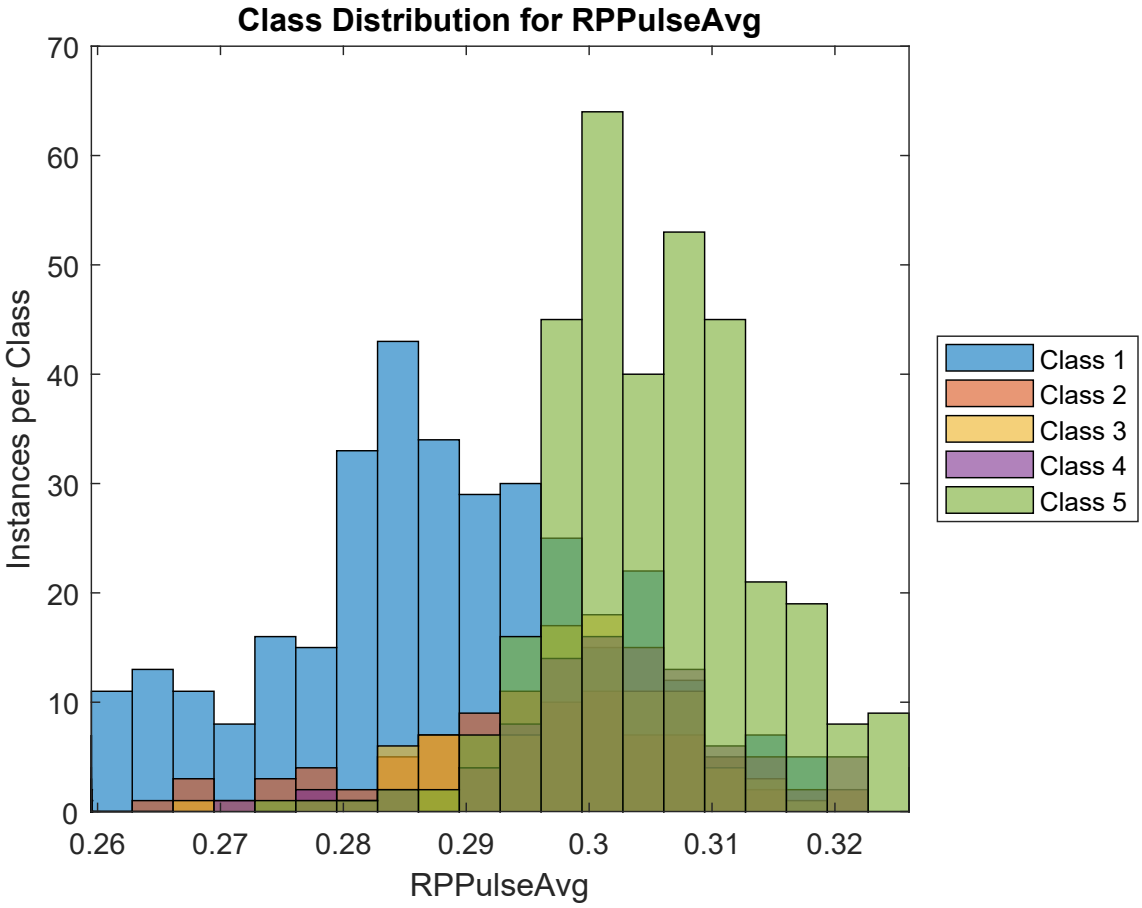

Figure 5.23: Histogram of the class distribution of the RPulseAvg feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.296	0.220	0.167	0.112
	Class 2	0.296	-	0.408	0.317	0.236
	Class 3	0.220	0.408	-	0.396	0.300
	Class 4	0.167	0.317	0.396	-	0.400
	Class 5	0.112	0.236	0.300	0.400	-

Table 5.19: Area under the curve for class versus class performance of the feature RPulseAvg.

The RPulseAvg feature is the ratio between the power of the Welch periodogram in the frequency range associated with good signal, 1-2.25Hz, and the frequency range from 0-8Hz, computed for the averaged pulse. This was computed according to Equation 5.10.

A histogram depicting the class distribution of the feature is shown in Figure 5.23. According to the histogram, the feature provides good discrimination discrimination between classes 1 and 5. Additionally, the table of AUC values in Table 5.19 indicates that the feature provides good discrimination between classes 1 and 4. Other class combinations yield poor or moderate class discrimination.

5.4.6 Skewness of Average Pulse (SPulseAvg)

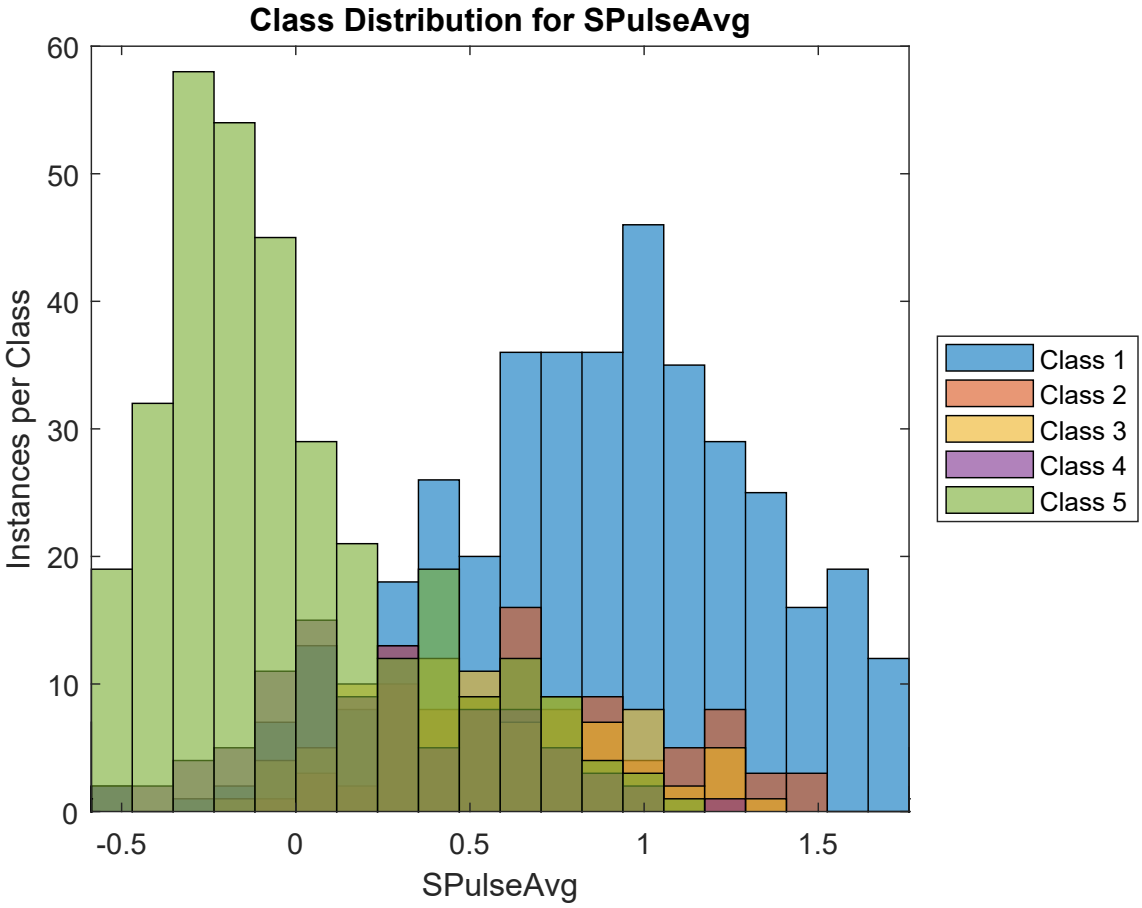

Figure 5.24: Histogram of the class distribution of the SPulseAvg feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.382	0.276	0.136	0.065
	Class 2	0.382	-	0.365	0.175	0.078
	Class 3	0.276	0.365	-	0.279	0.127
	Class 4	0.136	0.175	0.279	-	0.285
	Class 5	0.065	0.078	0.127	0.285	-

Table 5.20: Area under the curve for class versus class performance of the feature SPulseAvg.

The SPulseAvg feature is the skewness of the average pulse. It was implemented in Matlab using the *skewness* function. The skewness is a statistic measure of the symmetry of a distribution. Data segments belonging to lower quality classes were expected to asymmetrical distributions.

A histogram depicting the class distribution of the feature is shown in Figure 5.24. According to the histogram, the feature provides good discrimination discrimination between classes 1 and 5. Additionally, the table of AUC values in Table 5.20 indicates that the feature provides good discrimination of class 4 from classes 1 and 2, and of class 5 from classes 1, 2, and 3. Other class combinations yield poor or moderate discrimination.

5.5 Statistics of Features Per Pulse

Individual pulses within the 10 second data segment are identified using Billauer's algorithm for peak detection, implemented with the custom Matlab function *peakdet* [29]. This pulse segmentation procedure is identical to the process used in section 5.4. Each detected peak is assumed to represent the peak of a PPG pulse, as seen in Figure 5.13. The individual PPG pulses are segmented by taking a fixed window of 61 data points (approximately 0.95 seconds), with 18 data points before the peak, and 42 data points following the peak. This window size was manually optimized on a small segment of visually identified class 5 PPG data. The uneven window length on either side of the peak reflects the morphology of the ideal PPG pulse, which is characterized by a quick rise to the peak, followed by a slow fall. The metrics are computed for each individual PPG pulse, then statistics such as the mean, median and standard deviation are computed on those metrics. These statistics are then used as features.

5.5.1 Mean Entropy Per Pulse (meanE)

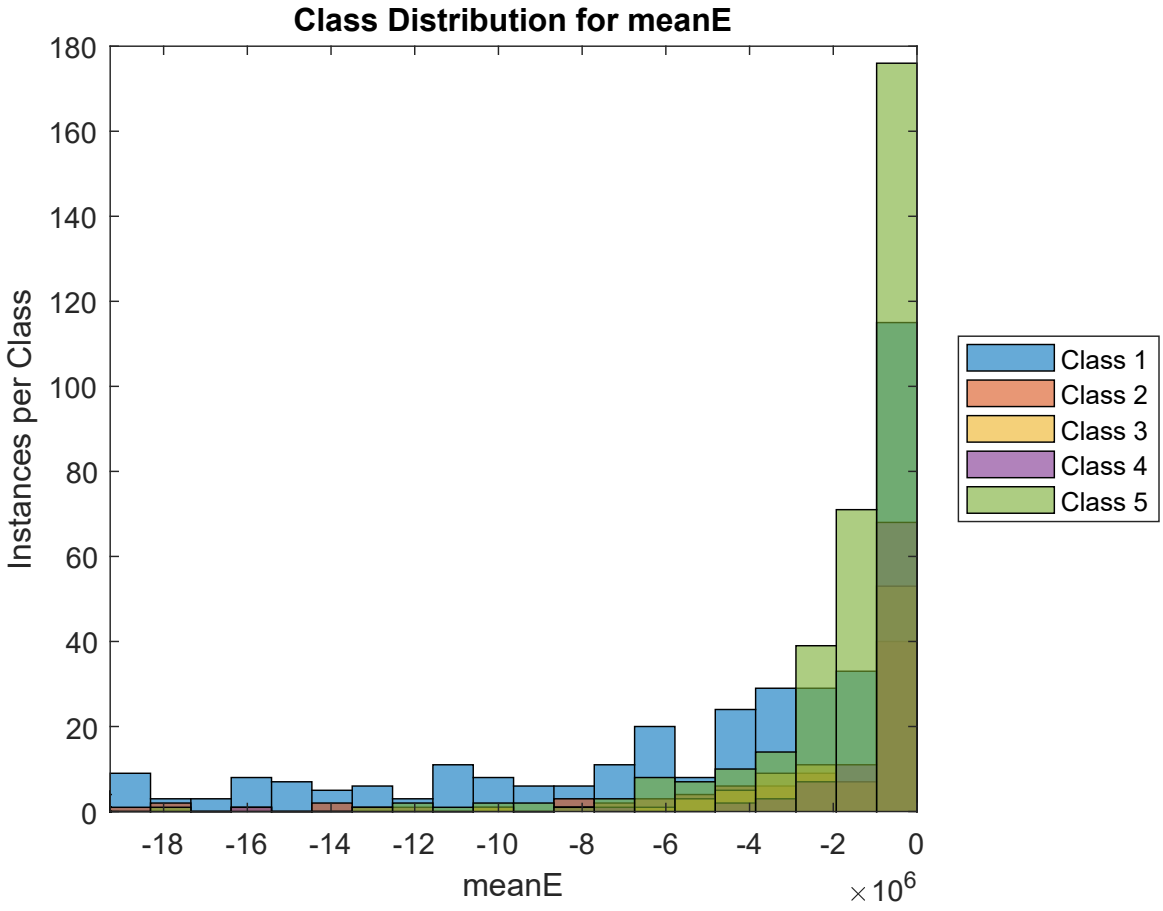

Figure 5.25: Histogram of the class distribution of the meanE feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.393	0.323	0.251	0.306
	Class 2	0.393	-	0.412	0.316	0.408
	Class 3	0.323	0.412	-	0.394	0.496
	Class 4	0.251	0.316	0.394	-	0.369
	Class 5	0.306	0.408	0.496	0.369	-

Table 5.21: Area under the curve for class versus class performance of the feature meanE.

The meanE feature is the mean of the entropy computed from each pulse in the PPG segment. Entropy is a measure of the uncertainty present in the PPG signal, using the difference between the probability density of the signal from a uniform distribution [28]. It was computed using Equation 5.3.

A histogram depicting the class distribution of the feature is shown in Figure 5.25. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.21, though the feature does provide moderate discrimination between classes 1 and 4.

5.5.2 Mean Kurtosis Per Pulse (meanK)

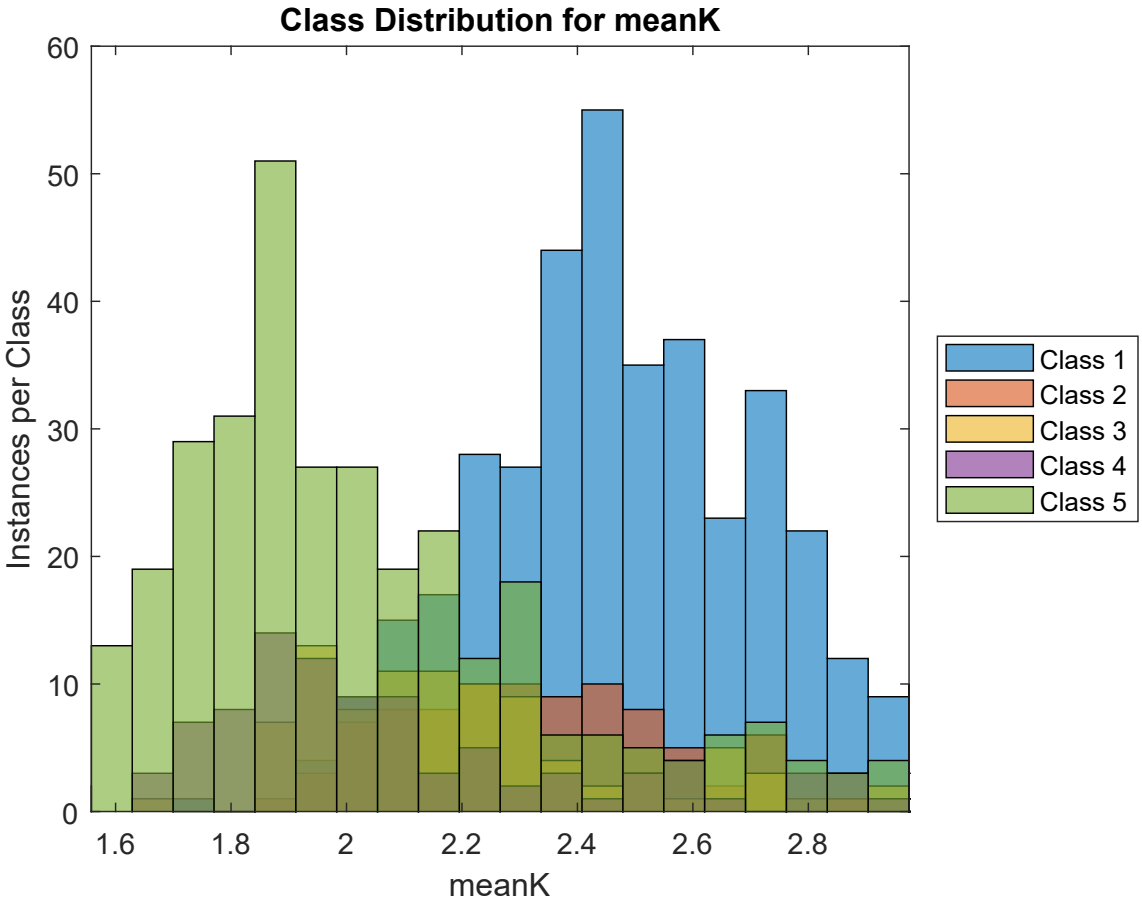

Figure 5.26: Histogram of the class distribution of the meanK feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.319	0.236	0.196	0.182
	Class 2	0.319	-	0.363	0.274	0.249
	Class 3	0.236	0.363	-	0.361	0.330
	Class 4	0.196	0.274	0.361	-	0.455
	Class 5	0.182	0.249	0.330	0.455	-

Table 5.22: Area under the curve for class versus class performance of the feature meanK.

The meanK feature is the mean of the kurtosis computed from each pulse in the PPG segment. Kurtosis represents the distribution of the data around the mean. It is believed that this will vary according to signal quality. It was implemented in Matlab using the *kurtosis* function.

A histogram depicting the class distribution of the feature is shown in Figure 5.26. According to the histogram, the feature provides good discrimination between classes 1 and 5. Additionally, according to the table of AUC values in Table 5.22, the feature also provides good discrimination between classes 1 and 4.

5.5.3 Mean Noise Ratio Per Pulse (meanN)

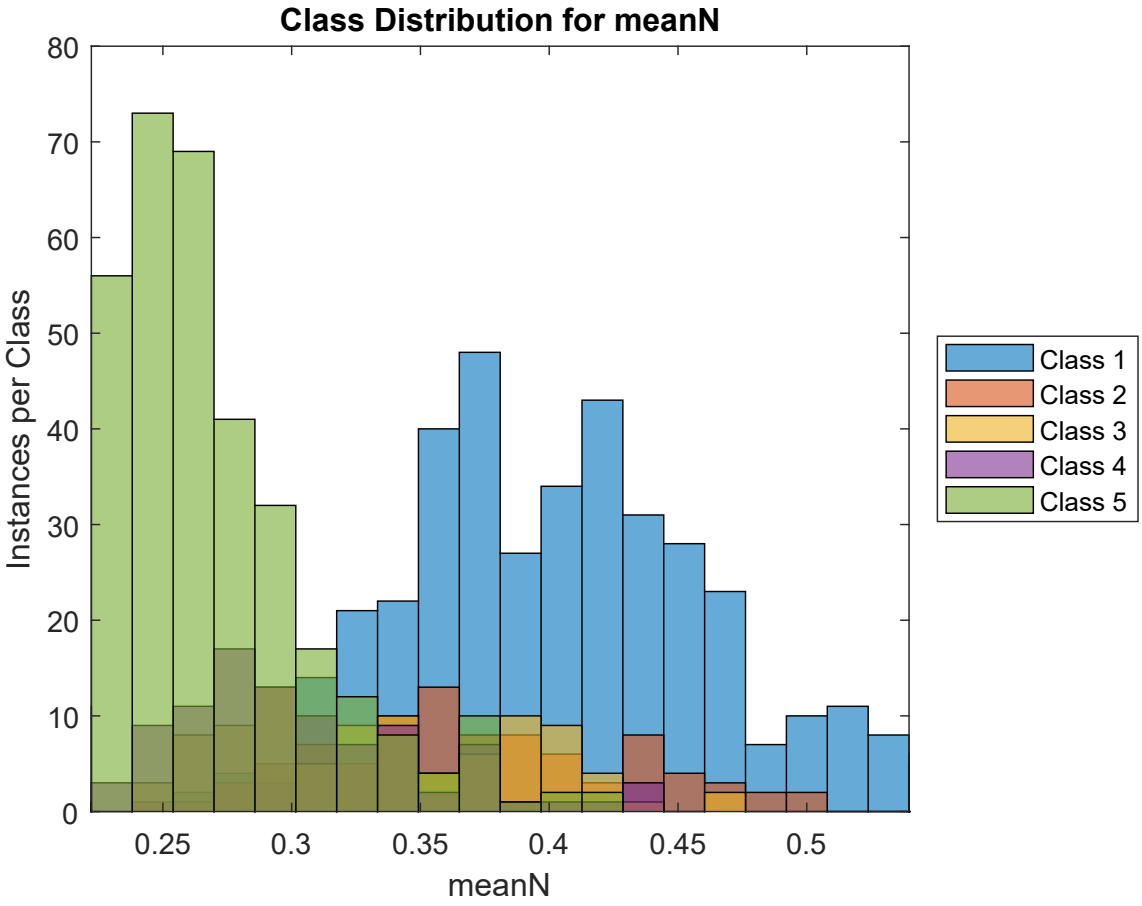

Figure 5.27: Histogram of the class distribution of the meanN feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.349	0.210	0.095	0.032
	Class 2	0.349	-	0.333	0.184	0.069
	Class 3	0.210	0.333	-	0.332	0.159
	Class 4	0.095	0.184	0.332	-	0.273
	Class 5	0.032	0.069	0.159	0.273	-

Table 5.23: Area under the curve for class versus class performance of the feature meanN.

The meanN feature is the mean of the signal-to-noise computed from each pulse in the PPG segment, using Equation 5.2. The implementation here differs from the implementation of the NoiseRatio feature, in that the signal is not filtered. The high-pass component of the band-pass filtering was deemed unnecessary as baseline drift was not expected to be present in the short window length of the PPG pulses. The low-pass component was also deemed unnecessary as the presence of high frequencies were expected to be a good indicator of low signal quality.

A histogram depicting the class distribution of the feature is shown in Figure 5.27. According to the histogram, the feature provides good discrimination between classes 1 and 5. Additionally, according to the table of AUC values in Table 5.23, the feature also provides good discrimination of class 4 from classes 1 and 2, and class 5 from classes 1, 2, and 3.

5.5.4 Mean Relative Power Per Pulse (meanR)

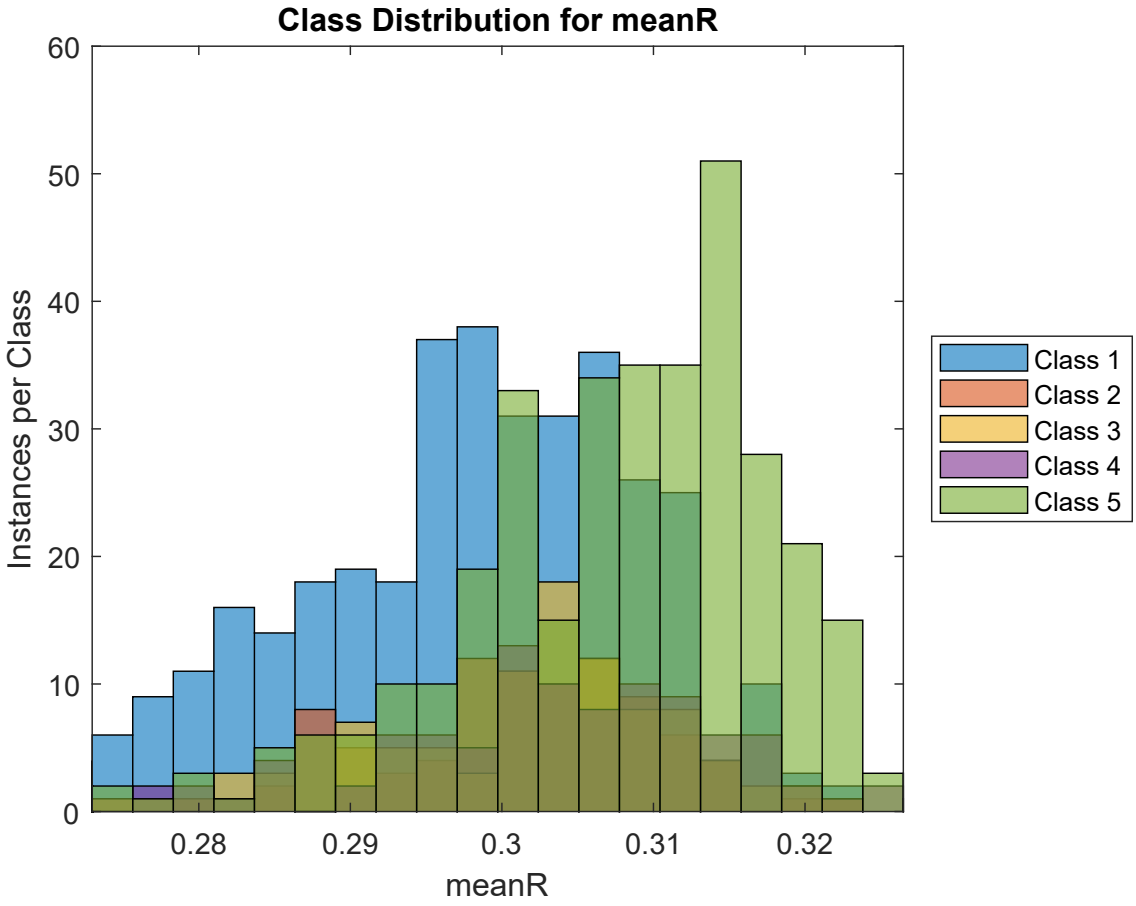

Figure 5.28: Histogram of the class distribution of the meanR feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.379	0.359	0.322	0.224
	Class 2	0.379	-	0.463	0.421	0.295
	Class 3	0.359	0.463	-	0.451	0.322
	Class 4	0.322	0.421	0.451	-	0.377
	Class 5	0.224	0.295	0.322	0.377	-

Table 5.24: Area under the curve for class versus class performance of the feature meanR.

The meanR feature is the mean of the relative power computed from each pulse in the PPG segment. This is the ratio between the power of the Welch periodogram in the frequency range associated with good signal, 1-2.25Hz, and the frequency range from 0-8Hz, computed for each pulse, then averaged. This was computed according to Equation 5.10.

A histogram depicting the class distribution of the feature is shown in Figure 5.28. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.24, though the feature does provide moderate discrimination of class 5 from classes 1 and 2.

5.5.5 Mean Skewness Per Pulse (meanS)

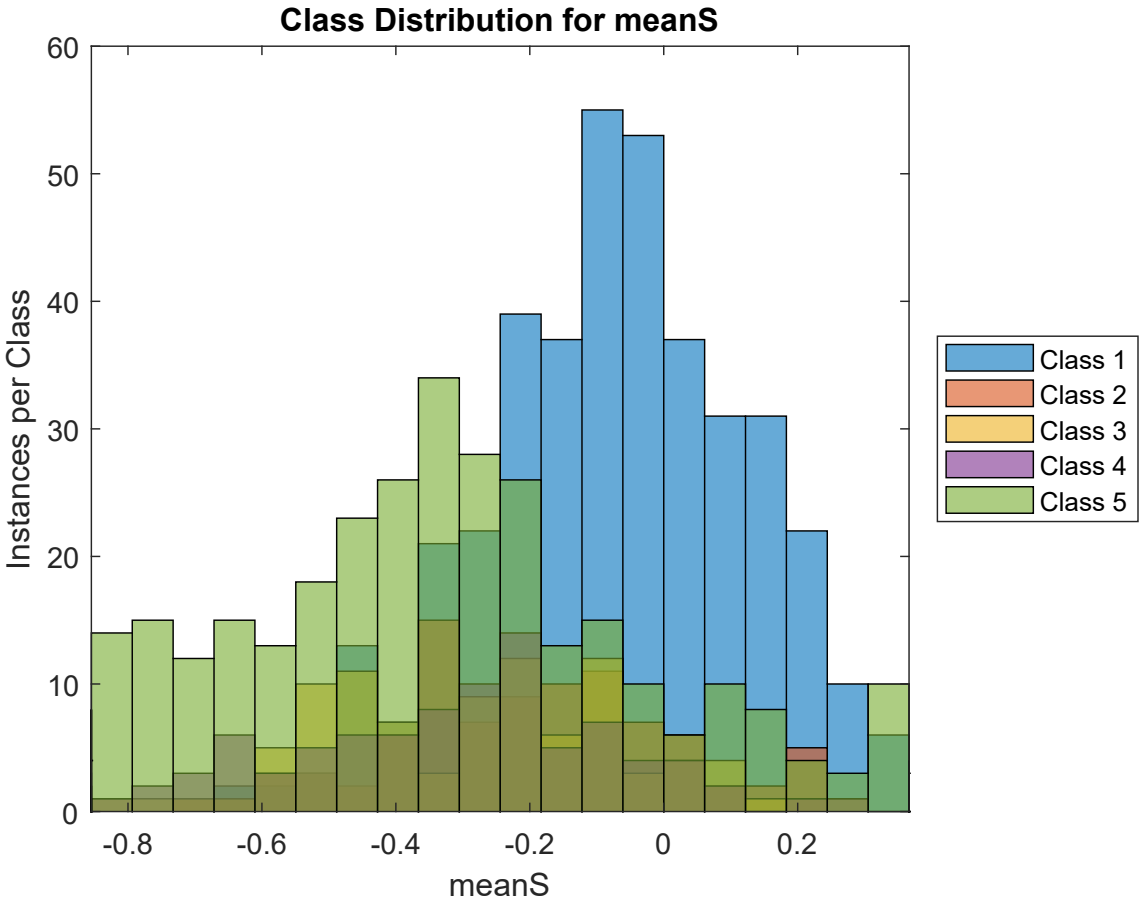

Figure 5.29: Histogram of the class distribution of the meanS feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.324	0.232	0.224	0.221
	Class 2	0.324	-	0.383	0.364	0.331
	Class 3	0.232	0.383	-	0.472	0.429
	Class 4	0.224	0.364	0.472	-	0.452
	Class 5	0.221	0.331	0.429	0.452	-

Table 5.25: Area under the curve for class versus class performance of the feature meanS.

The meanS feature is the mean of the skewness computed from each pulse in the PPG segment. It was implemented in Matlab using the *skewness* function. The skewness is a statistic measure of the symmetry of a distribution. Data segments belonging to lower quality classes were expected to asymmetrical distributions.

A histogram depicting the class distribution of the feature is shown in Figure 5.29. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.25, though the feature does provide moderate discrimination of class 1 from classes 3, 4, and 5.

5.5.6 Median Entropy Per Pulse (medianE)

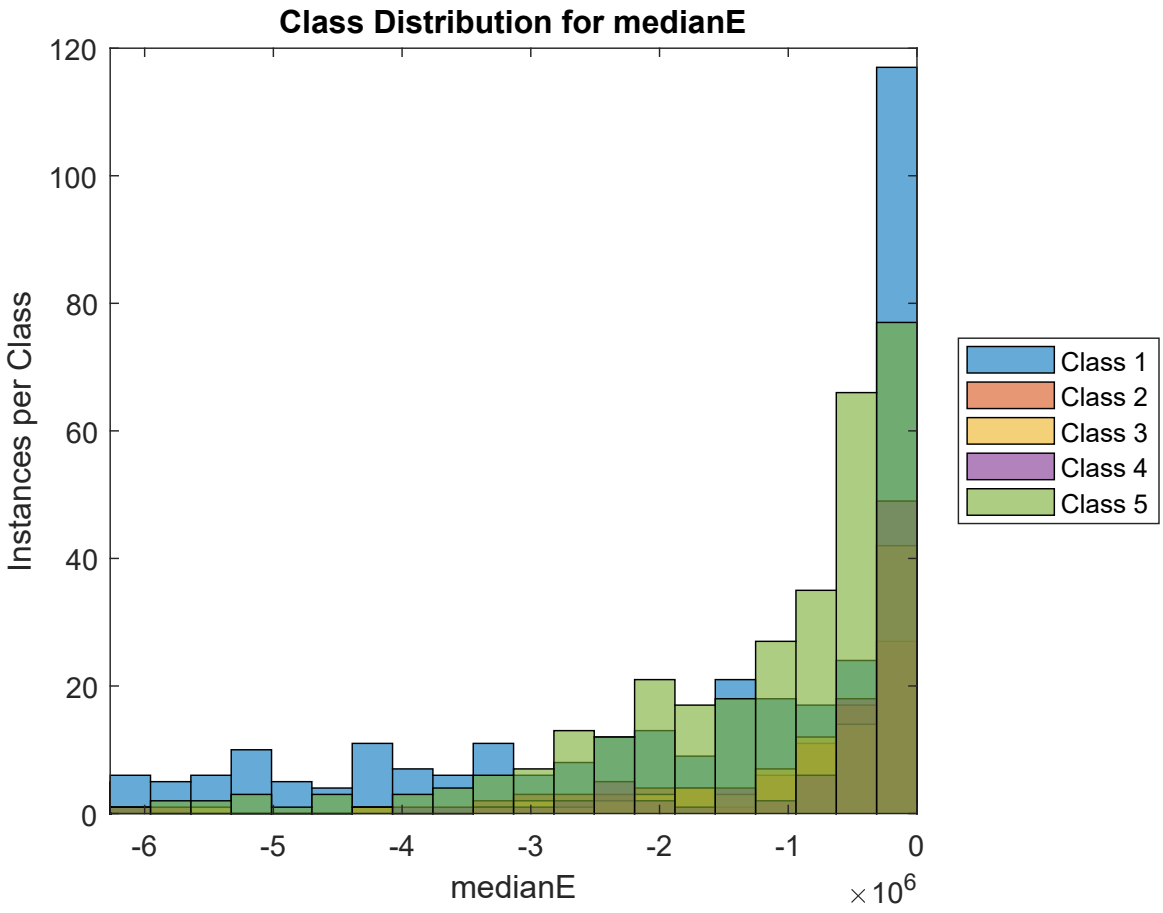

Figure 5.30: Histogram of the class distribution of the medianE feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.394	0.336	0.318	0.424
	Class 2	0.394	-	0.421	0.389	0.444
	Class 3	0.336	0.421	-	0.464	0.354
	Class 4	0.318	0.389	0.464	-	0.319
	Class 5	0.424	0.444	0.354	0.319	-

Table 5.26: Area under the curve for class versus class performance of the feature medianE.

The medianE feature is the median of the entropy computed from each pulse in the PPG segment. Entropy is a measure of the uncertainty present in the PPG signal, using the difference between the probability density of the signal from a uniform distribution [28]. It was computed using Equation 5.3.

A histogram depicting the class distribution of the feature is shown in Figure 5.30. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.26.

5.5.7 Median Kurtosis Per Pulse (medianK)

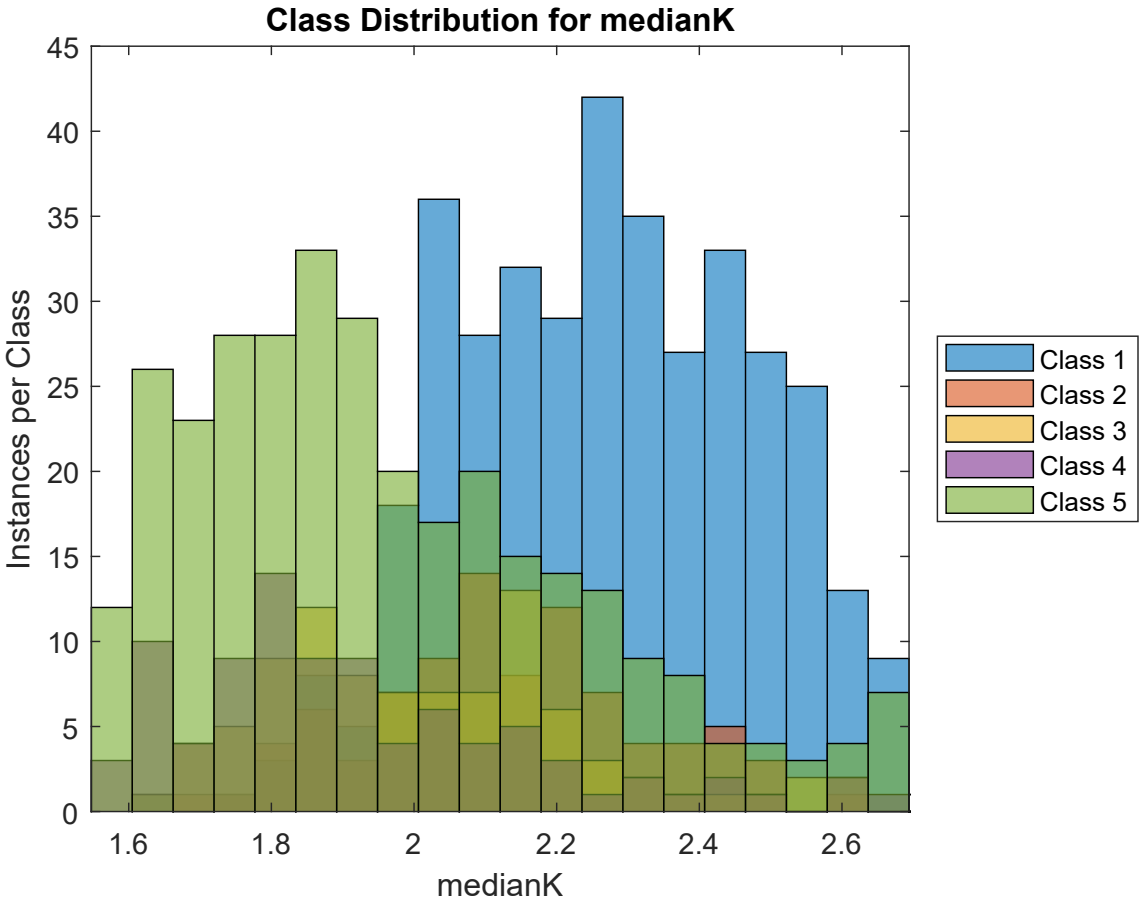

Figure 5.31: Histogram of the class distribution of the medianK feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.305	0.231	0.183	0.224
	Class 2	0.305	-	0.372	0.274	0.329
	Class 3	0.231	0.372	-	0.372	0.428
	Class 4	0.183	0.274	0.372	-	0.457
	Class 5	0.224	0.329	0.428	0.457	-

Table 5.27: Area under the curve for class versus class performance of the feature medianK.

The medianK feature is the median of the kurtosis computed from each pulse in the PPG segment. Kurtosis represents the distribution of the data around the mean. It is believed that this will vary according to signal quality. It was implemented in Matlab using the *kurtosis* function.

A histogram depicting the class distribution of the feature is shown in Figure 5.31. According to the histogram, the feature provides moderate discrimination between classes 1 and 5. Additionally, according to the table of AUC values in Table 5.27, the feature also provides good discrimination between classes 1 and 4.

5.5.8 Median Noise Ratio Per Pulse (medianN)

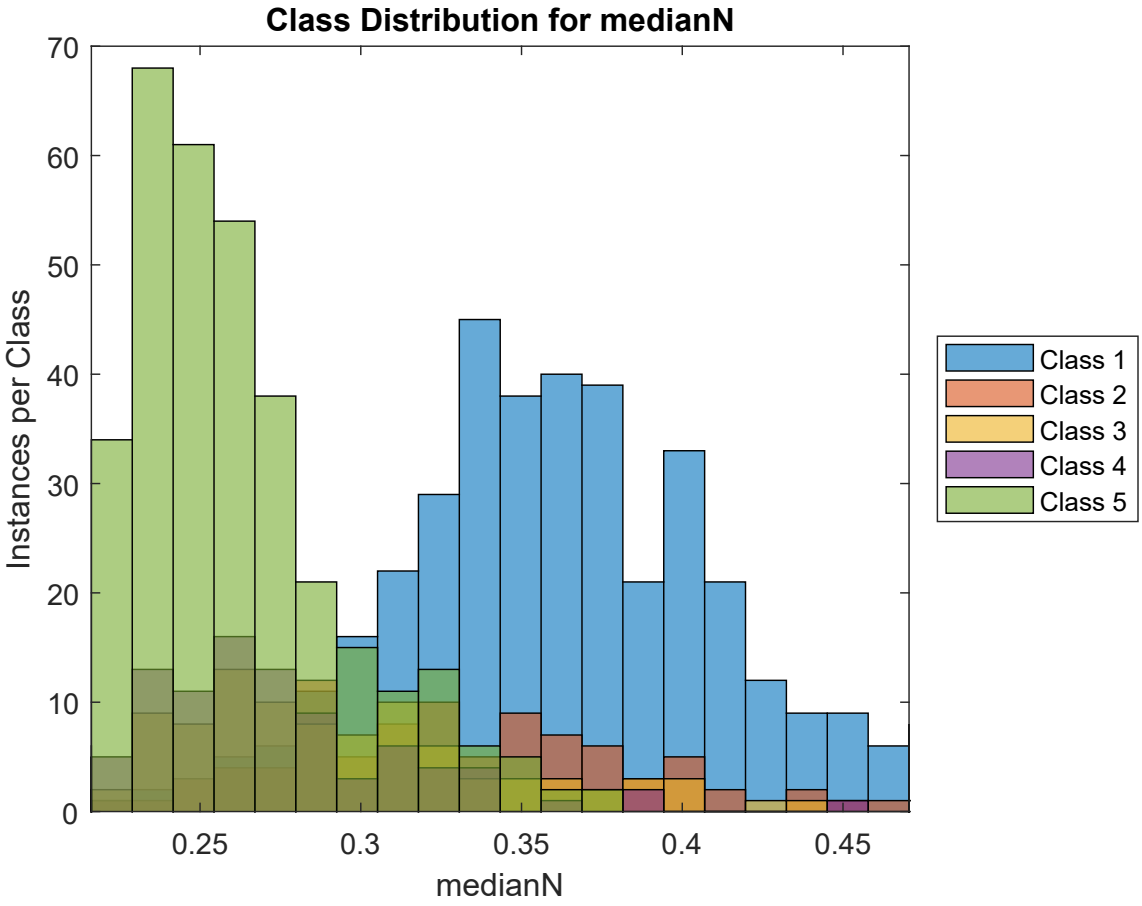

Figure 5.32: Histogram of the class distribution of the medianN feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.309	0.150	0.073	0.046
	Class 2	0.309	-	0.288	0.169	0.124
	Class 3	0.150	0.288	-	0.358	0.300
	Class 4	0.073	0.169	0.358	-	0.440
	Class 5	0.046	0.124	0.300	0.440	-

Table 5.28: Area under the curve for class versus class performance of the feature medianN.

The medianN feature is the median of the signal-to-noise computed from each pulse in the PPG segment, using Equation 5.2. The implementation here differs from the implementation of the NoiseRatio feature, in that the signal is not filtered. The high-pass component of the band-pass filtering was deemed unnecessary as baseline drift was not expected to be present in the short window length of the PPG pulses. The low-pass component was also deemed unnecessary as the presence of high frequencies were expected to be a good indicator of low signal quality.

A histogram depicting the class distribution of the feature is shown in Figure 5.32. According to the histogram, the feature provides good discrimination between classes 1 and 5. Additionally, according to the table of AUC values in Table 5.28, the feature also provides good discrimination of class 1 from classes 3, 4 and 5, and class 5 from classes 1 and 2.

5.5.9 Median Relative Power Per Pulse (medianR)

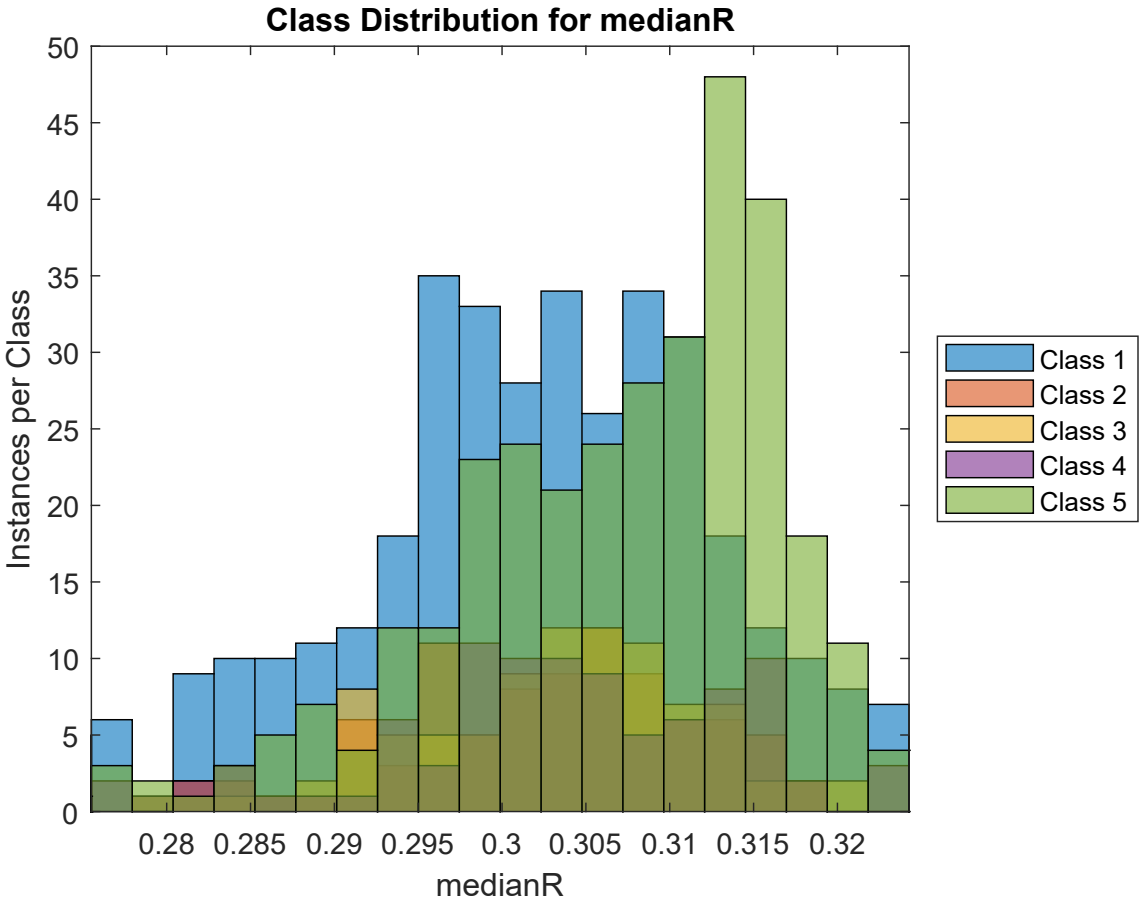

Figure 5.33: Histogram of the class distribution of the medianR feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.435	0.426	0.388	0.305
	Class 2	0.435	-	0.488	0.446	0.344
	Class 3	0.426	0.488	-	0.459	0.351
	Class 4	0.388	0.446	0.459	-	0.402
	Class 5	0.305	0.344	0.351	0.402	-

Table 5.29: Area under the curve for class versus class performance of the feature medianR.

The medianR feature is the median of the relative power computed from each pulse in the PPG segment. This is the ratio between the power of the Welch periodogram in the frequency range associated with good signal, 1-2.25Hz, and the frequency range from 0-8Hz, computed for each pulse, then the median of those values is taken as a feature. The relative power was computed according to Equation 5.10.

A histogram depicting the class distribution of the feature is shown in Figure 5.33. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.29.

5.5.10 Median Skewness Per Pulse (medianS)

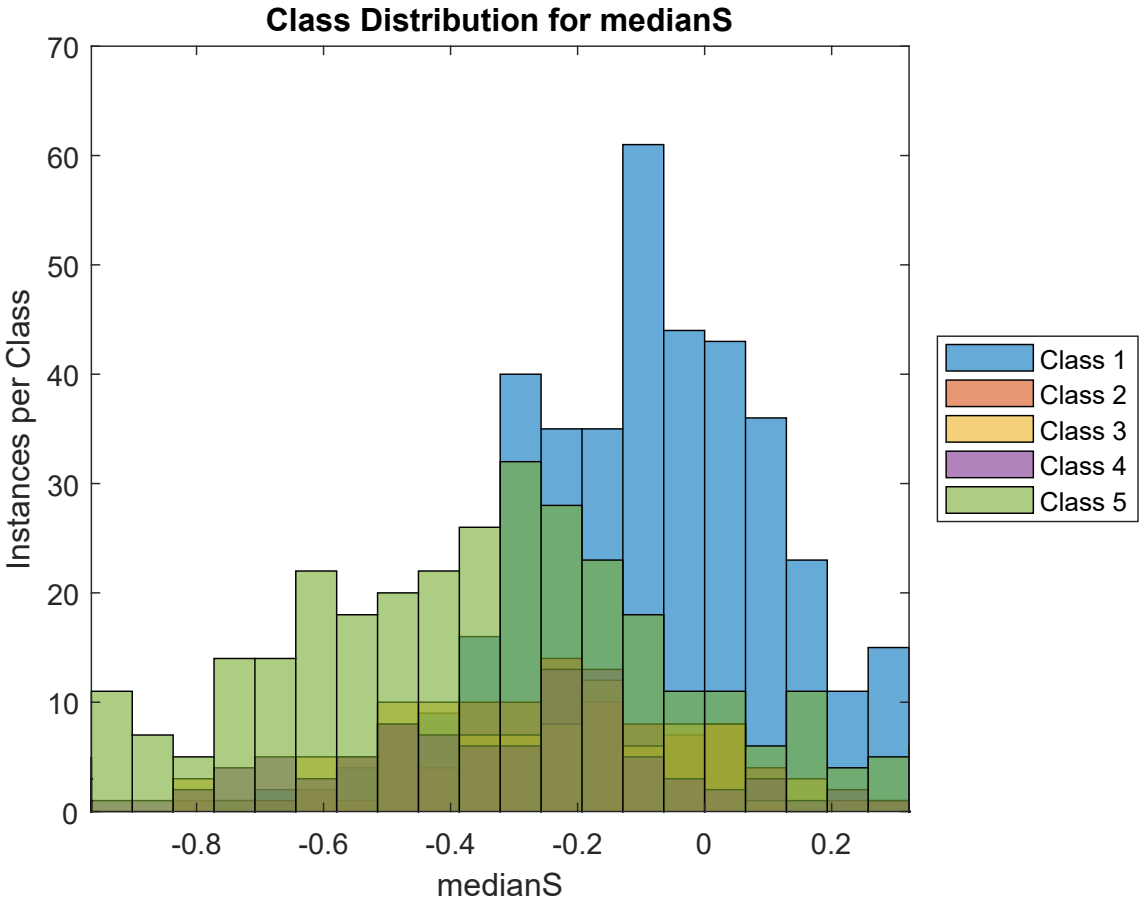

Figure 5.34: Histogram of the class distribution of the medianS feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.319	0.263	0.246	0.253
	Class 2	0.319	-	0.419	0.387	0.374
	Class 3	0.263	0.419	-	0.459	0.443
	Class 4	0.246	0.387	0.459	-	0.477
	Class 5	0.253	0.374	0.443	0.477	-

Table 5.30: Area under the curve for class versus class performance of the feature medianS.

The medianS feature is the median of the skewness computed from each pulse in the PPG segment. It was implemented in Matlab using the *skewness* function. The skewness is a statistic measure of the symmetry of a distribution. Data segments belonging to lower quality classes were expected to asymmetrical distributions.

A histogram depicting the class distribution of the feature is shown in Figure 5.34. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.30, though the feature does provide moderate discrimination of class 1 from classes 3, 4, and 5.

5.5.11 Standard Deviation of Entropy Per Pulse (stdevE)

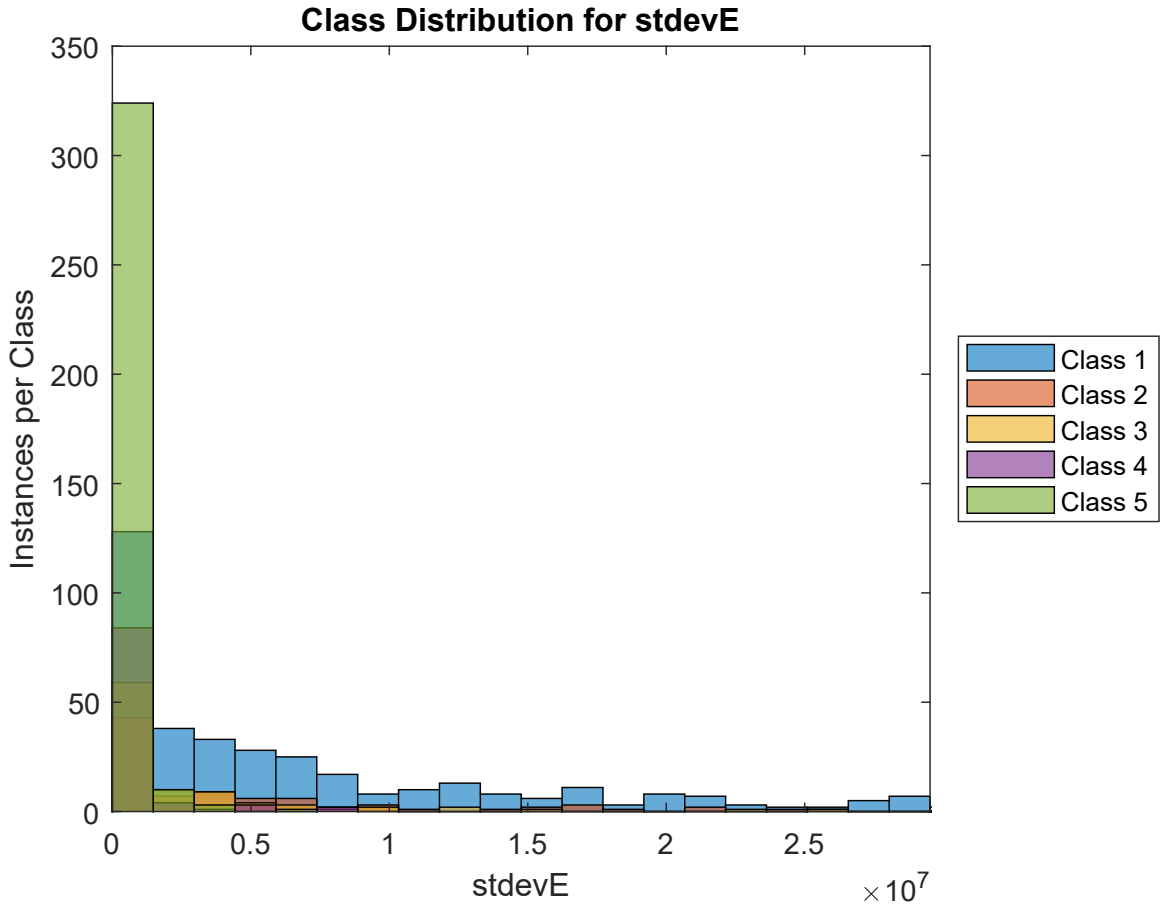

Figure 5.35: Histogram of the class distribution of the stdevE feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.400	0.316	0.203	0.161
	Class 2	0.400	-	0.408	0.262	0.207
	Class 3	0.316	0.408	-	0.334	0.275
	Class 4	0.203	0.262	0.334	-	0.451
	Class 5	0.161	0.207	0.275	0.451	-

Table 5.31: Area under the curve for class versus class performance of the feature stdevE.

The stdevE feature is the standard deviation of the entropy computed from each pulse in the PPG segment. Entropy is a measure of the uncertainty present in the PPG signal, using the difference between the probability density of the signal from a uniform distribution [28]. It was computed using Equation 5.3.

A histogram depicting the class distribution of the feature is shown in Figure 5.35. According to the histogram, the feature provides poor discriminability between classes. However, the table of AUC values in Table 5.31 shows that the feature provides good discrimination between classes 1 and 5.

5.5.12 Standard Deviation of Kurtosis Per Pulse (stdevK)

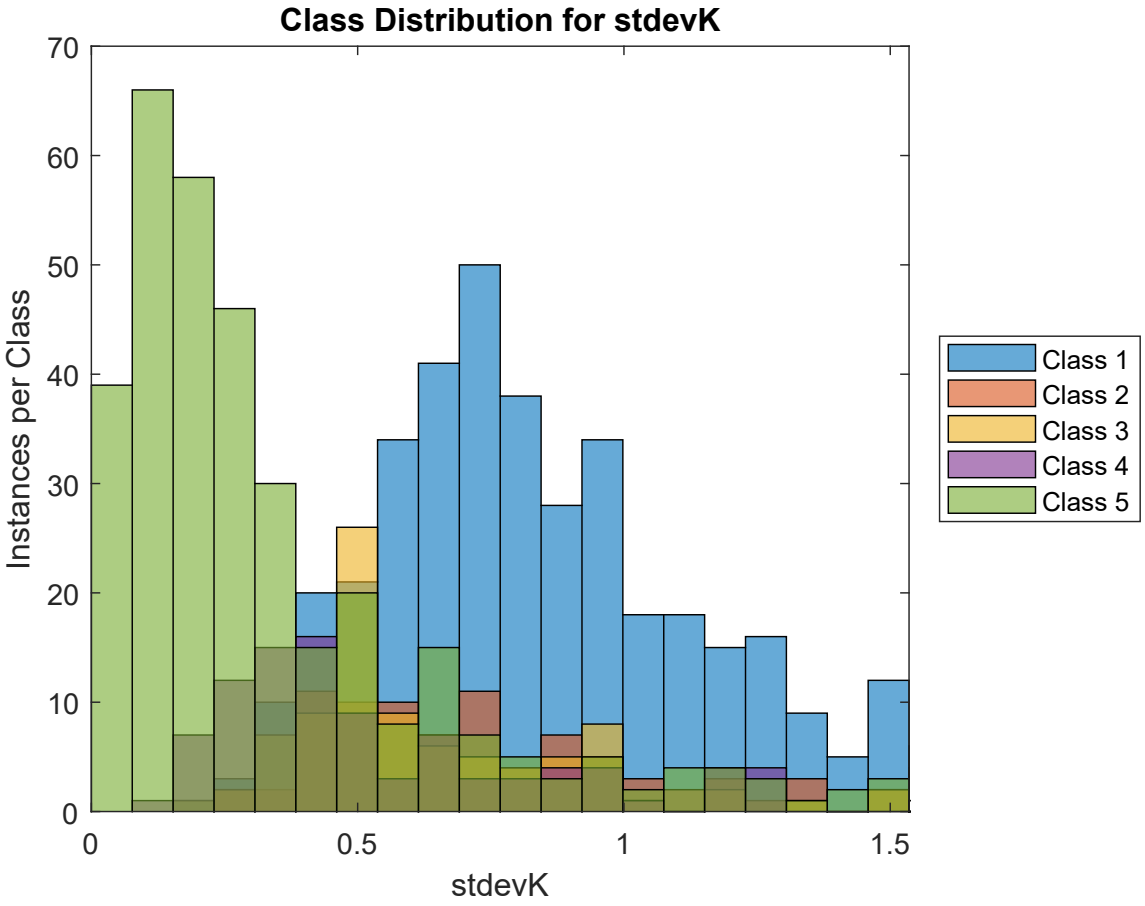

Figure 5.36: Histogram of the class distribution of the stdevK feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.418	0.326	0.232	0.112
	Class 2	0.418	-	0.408	0.284	0.138
	Class 3	0.326	0.408	-	0.345	0.172
	Class 4	0.232	0.284	0.345	-	0.255
	Class 5	0.112	0.138	0.172	0.255	-

Table 5.32: Area under the curve for class versus class performance of the feature stdevK.

The stdevK feature is the standard deviation of the kurtosis computed from each pulse in the PPG segment. Kurtosis represents the distribution of the data around the mean. It is believed that this will vary according to signal quality. It was implemented in Matlab using the *kurtosis* function.

A histogram depicting the class distribution of the feature is shown in Figure 5.36. According to the histogram, the feature provides moderate discrimination between classes 1 and 5. Additionally, according to the table of AUC values in Table 5.32, the feature also provides good discrimination of class 5 from classes 1, 2, and 3.

5.5.13 Standard Deviation of Noise Ratio Per Pulse (stdevN)

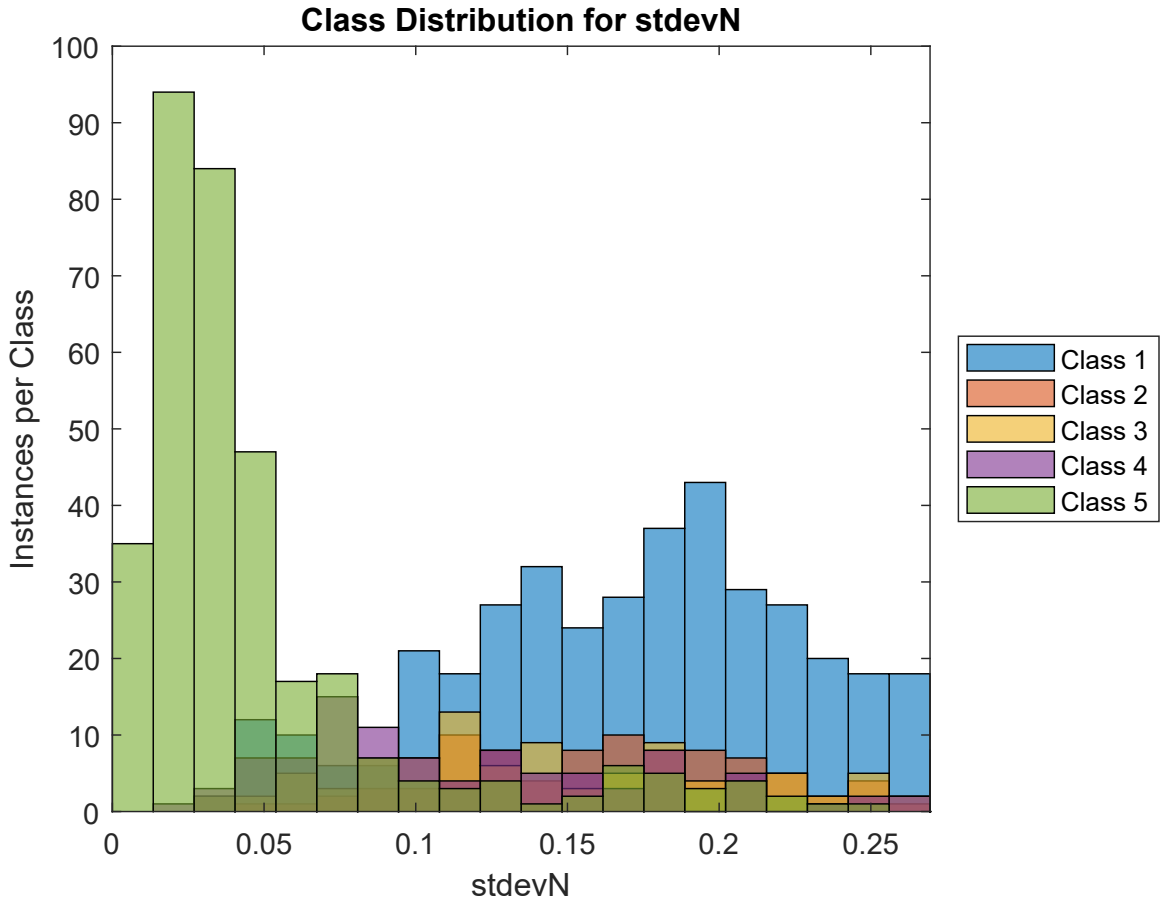

Figure 5.37: Histogram of the class distribution of the stdevN feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.464	0.370	0.246	0.058
	Class 2	0.464	-	0.395	0.254	0.056
	Class 3	0.370	0.395	-	0.352	0.082
	Class 4	0.246	0.254	0.352	-	0.130
	Class 5	0.058	0.056	0.082	0.130	-

Table 5.33: Area under the curve for class versus class performance of the feature stdevN.

The stdevN feature is the standard deviation of the signal-to-noise computed from each pulse in the PPG segment, using Equation 5.2. The implementation here differs from the implementation of the NoiseRatio feature, in that the signal is not filtered. The high-pass component of the band-pass filtering was deemed unnecessary as baseline drift was not expected to be present in the short window length of the PPG pulses. The low-pass component was also deemed unnecessary as the presence of high frequencies were expected to be a good indicator of low signal quality.

A histogram depicting the class distribution of the feature is shown in Figure 5.37. According to the histogram, the feature provides good discrimination of class 5 from the other classes. This is corroborated by the table of AUC values in Table 5.33.

5.5.14 Standard Deviation of Relative Power Per Pulse (stdevR)

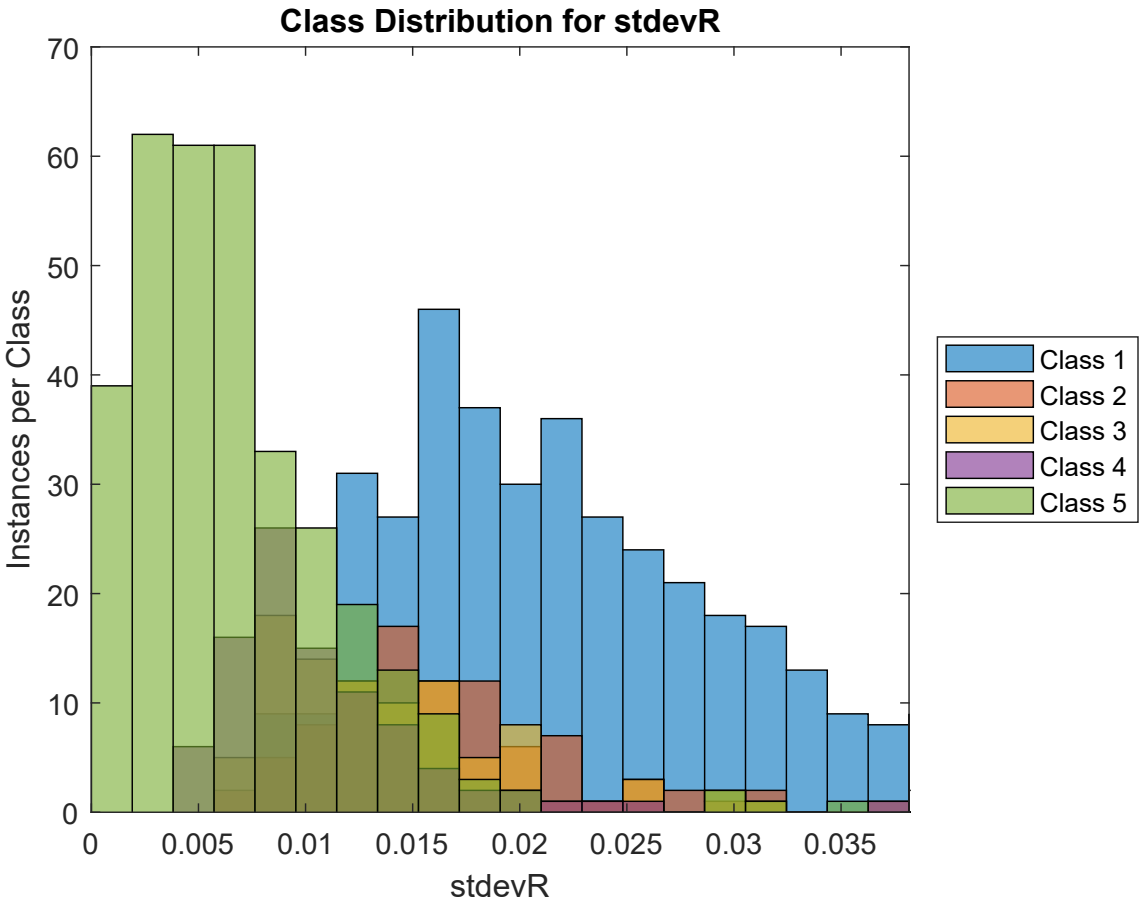

Figure 5.38: Histogram of the class distribution of the stdevR feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.263	0.209	0.094	0.053
	Class 2	0.263	-	0.382	0.191	0.096
	Class 3	0.209	0.382	-	0.298	0.147
	Class 4	0.094	0.191	0.298	-	0.251
	Class 5	0.053	0.096	0.147	0.251	-

Table 5.34: Area under the curve for class versus class performance of the feature stdevR.

The stdevR feature is the standard deviation of the relative power computed from each pulse in the PPG segment. This is the ratio between the power of the Welch periodogram in the frequency range associated with good signal, 1-2.25Hz, and the frequency range from 0-8Hz, computed for each pulse, then the standard deviation of those values is taken as a feature. The relative power was computed according to Equation 5.10.

A histogram depicting the class distribution of the feature is shown in Figure 5.38. According to the histogram, the feature provides good discrimination between classes 1 and 5. Additionally, the table of AUC values in Table 5.34 indicates good or moderate discrimination for all class combinations, with the exception of the class 2 and 3 combination.

5.5.15 Standard Deviation of Skewness Per Pulse (stdevS)

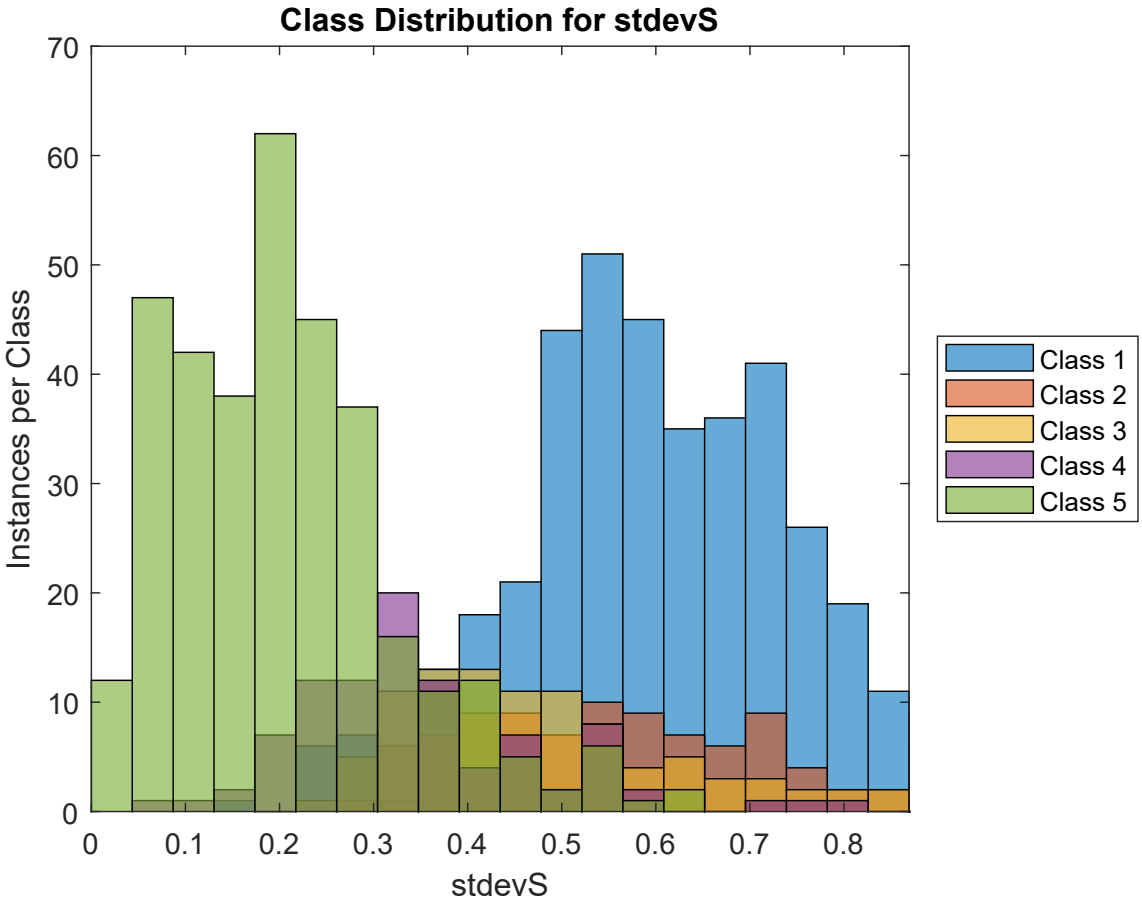

Figure 5.39: Histogram of the class distribution of the stdevS feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.396	0.256	0.125	0.025
	Class 2	0.396	-	0.350	0.172	0.035
	Class 3	0.256	0.350	-	0.262	0.061
	Class 4	0.125	0.172	0.262	-	0.184
	Class 5	0.025	0.035	0.061	0.184	-

Table 5.35: Area under the curve for class versus class performance of the feature stdevS.

The stdevS feature is the standard deviation of the skewness computed from each pulse in the PPG segment. It was implemented in Matlab using the *skewness* function. The skewness is a statistic measure of the symmetry of a distribution. Data segments belonging to lower quality classes were expected to asymmetrical distributions.

A histogram depicting the class distribution of the feature is shown in Figure 5.39. According to the histogram, the feature provides good discrimination between classes 1 and 5. According to the table of AUC values in Table 5.35, the feature also provides good discrimination of class 4 from classes 1, 2, and 5, and of class class 5 from all other classes.

5.5.16 Mean Signal to Noise Ratio Per Pulse (meanSNR)

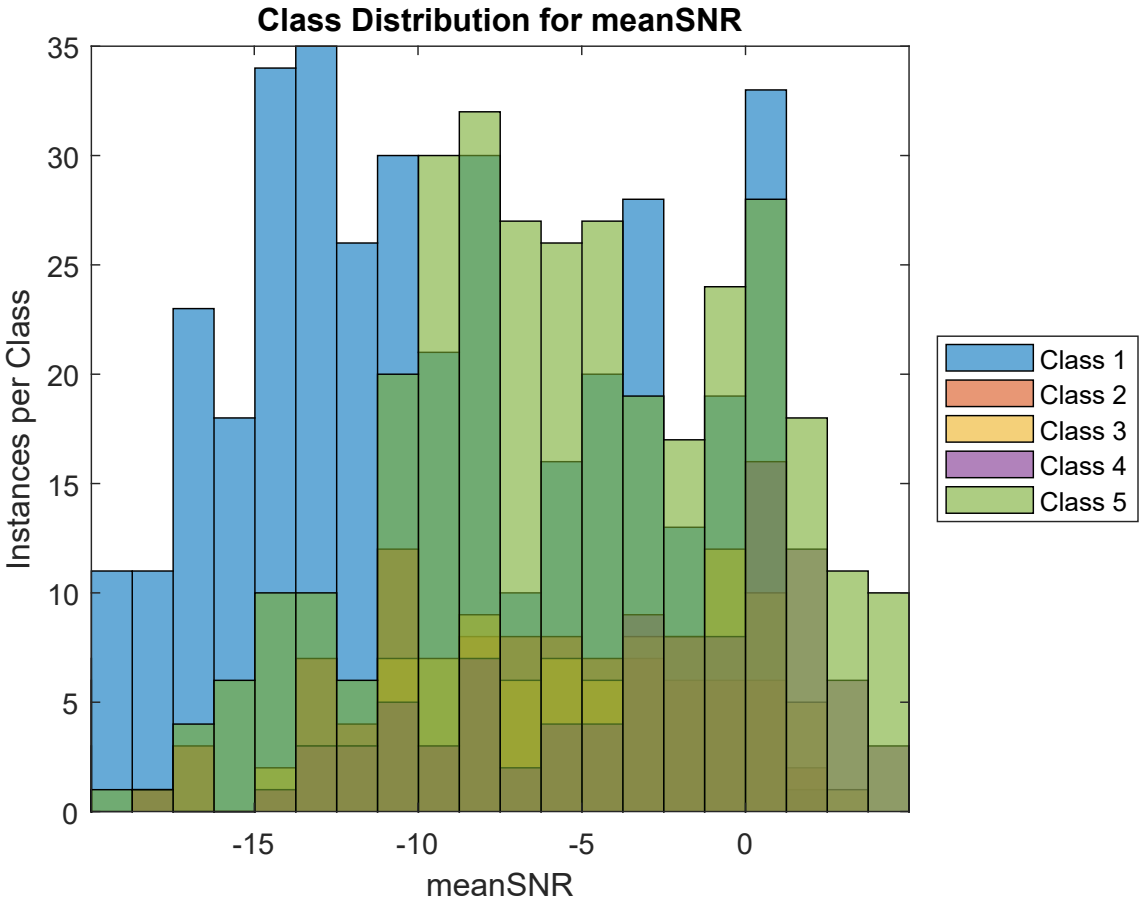

Figure 5.40: Histogram of the class distribution of the meanSNR feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.369	0.280	0.200	0.291
	Class 2	0.369	-	0.391	0.280	0.403
	Class 3	0.280	0.391	-	0.368	0.498
	Class 4	0.200	0.280	0.368	-	0.388
	Class 5	0.291	0.403	0.498	0.388	-

Table 5.36: Area under the curve for class versus class performance of the feature meanSNR.

The meanSNR feature is the mean of the signal-to-noise ratios (SNR) computed for each PPG pulse in the 10 second segment. The SNR is computed differently that the NoiseRatio feature. Equation 5.7 is used to obtain the SNR, where the signal is the ideal pulse template, and the noise is the deviation of the PPG pulse from the template. The template was created by visual identification of 10 high quality pulses, which were then averaged.

$$SNR = 10 \times log_{10} \left[\frac{\frac{1}{N} \sum_{n=1}^{N} T[n]^2}{\frac{1}{N} \sum_{n=1}^{N} (x_{pulse}[n] - T[n])^2} \right] \tag{5.7}$$

Where N is the number of data points in each pulse, x_{pulse} is the PPG pulse, and T is the template pulse. This implementation of SNR was inspired by its usage in [33].

A histogram depicting the class distribution of the feature is shown in Figure 5.40. According to the histogram, the feature provides poor discrimination between classes. This is corroborated by the table of AUC values in Table 5.36, though some class combinations offer moderate discrimination.

5.5.17 Median Signal to Noise Ratio Per Pulse (medianSNR)

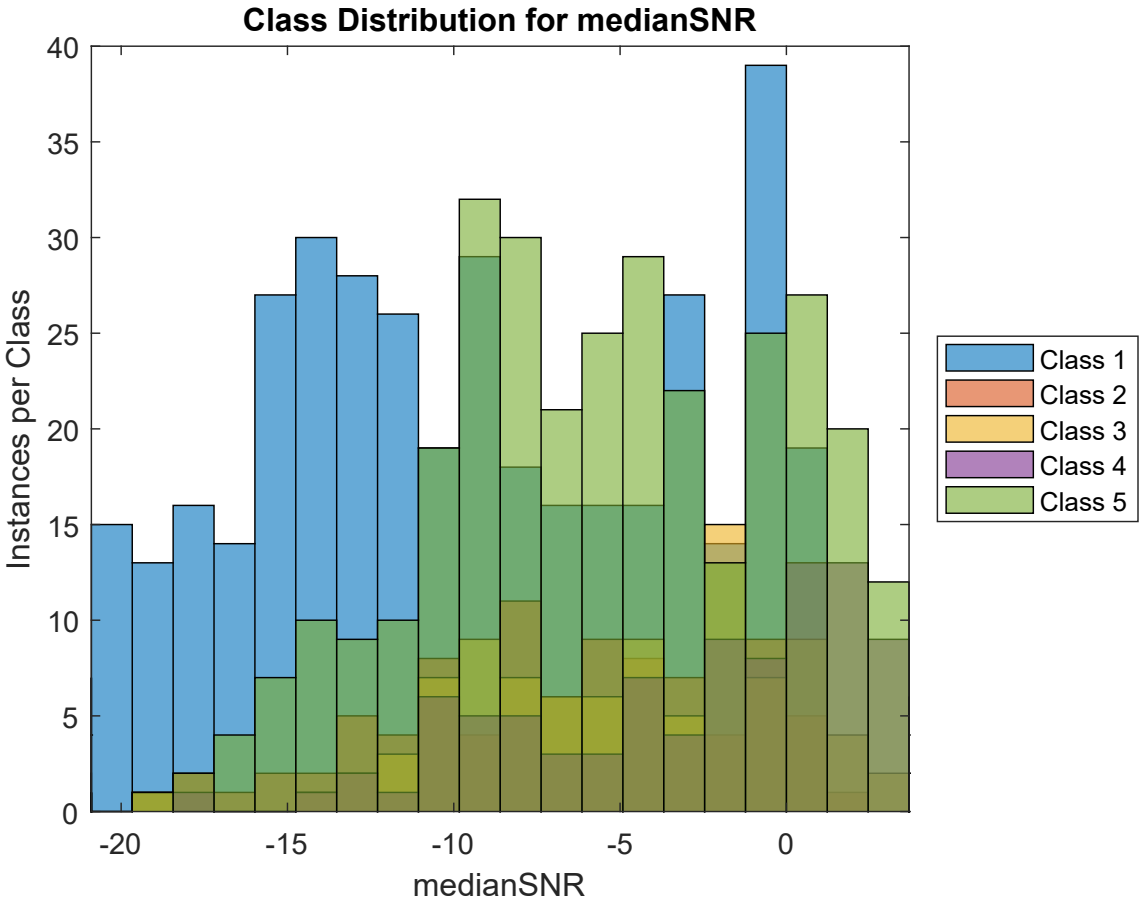

Figure 5.41: Histogram of the class distribution of the medianSNR feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.373	0.282	0.191	0.293
	Class 2	0.373	-	0.383	0.264	0.401
	Class 3	0.282	0.383	-	0.357	0.495
	Class 4	0.191	0.264	0.357	-	0.372
	Class 5	0.293	0.401	0.495	0.372	-

Table 5.37: Area under the curve for class versus class performance of the feature medianSNR.

The medianSNR feature is the median of the SNR computed for each PPG pulse in the 10 second segment. The SNR is computed differently that the NoiseRatio feature. Equation 5.7 is used to obtain the SNR, where the signal is the ideal pulse template, and the noise is the deviation of the PPG pulse from the template. The template was created by visual identification of 10 high quality pulses, which were then averaged.

A histogram depicting the class distribution of the feature is shown in Figure 5.41. According to the histogram, the feature provides poor discrimination between classes. However, the table of AUC values in Table 5.37 indicates good discrimination between classes 3 and 4. The feature provides poor or moderate discrimination for other class combinations.

5.6 Correlogram Features

It was postulated that there would be a difference in the periodicity of the PPG signals of different classes. Noise corruption due to random, non-periodic motions are expected to result in non-periodic motion artifacts in the PPG signal. Thus, an assessment of the periodicity of the PPG signal may assist in class discrimination. This was done using a correlogram.

A correlogram displays the autocorrelation of the PPG signal with itself on the y-axis, with corresponding time lags on the x-axis. Autocorrelation values could range from +1.0, indicating perfect correlation, to zero, indicating no correlation, to -1.0, indicating a perfect inverse correlation. Correlograms were constructed for each data segment. The autocorrelation of the PPG data was determined at lags of up to 3 seconds.

In the correlograms for each data segment, the first three peaks are identified using the *findpeaks* function in Matlab. The autocorrelation and location (in time) at each of the peaks are computed, and used as features. These features were inspired by the work of Wander and Morris [21], and previously used in preliminary work done by our research team [2].

5.6.1 Location of First Autocorrelation Peak (ACPeakLocs1)

The ACPeakLocs1 feature is the location of the first peak in the correlogram for each data segment. The location is represented as the time lag at which the peak occurs. A histogram depicting the class distribution of the feature is shown in Figure 5.42. According to the histogram, the feature does not provide any noticeable discriminability between classes 2,3, and 4. The feature does provide some discriminability between classes 1 and 5, though this separation is moderate. This is corroborated by the table

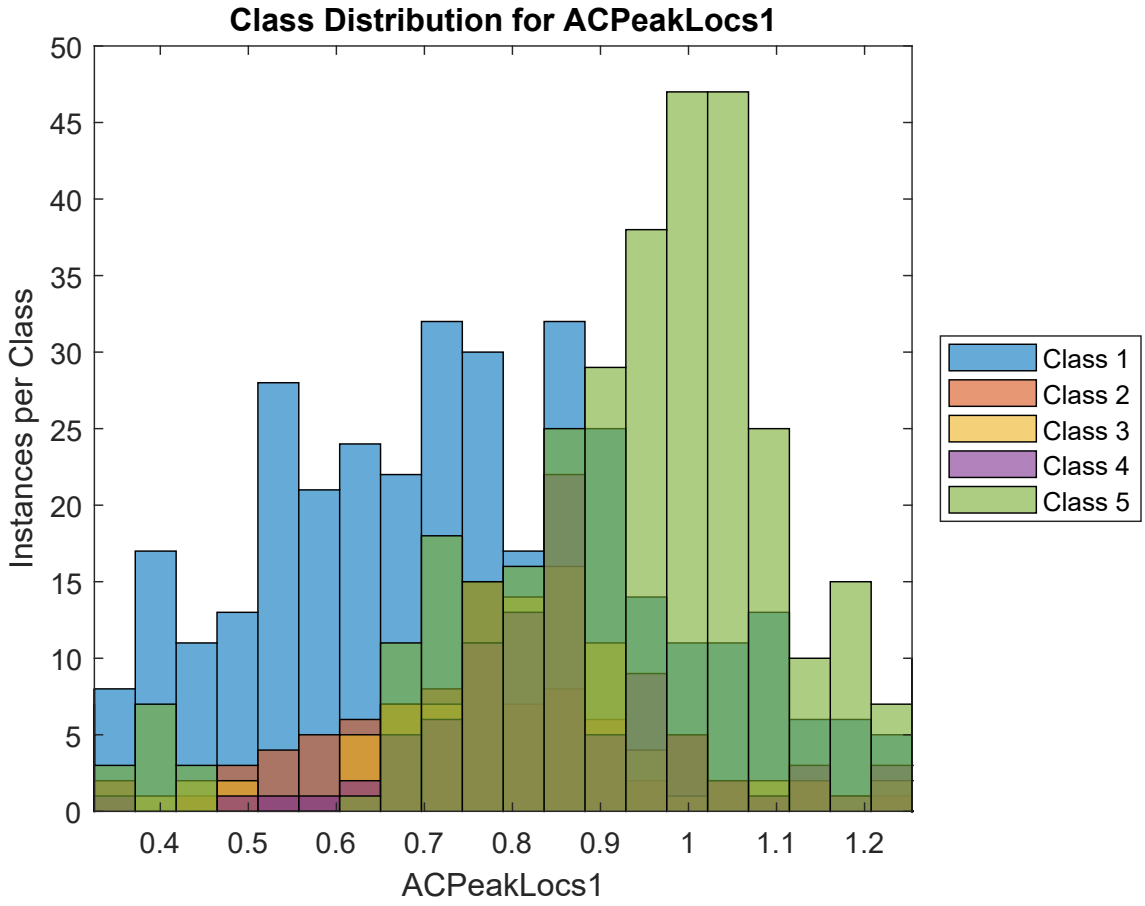

Figure 5.42: Histogram of the class distribution of the ACPeakLocs1 feature.

of AUC values in Table 5.38. Furthermore, the table indicates that there is moderate separation between class 5 from classes 2 and 3.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.438	0.385	0.347	0.252
	Class 2	0.438	-	0.422	0.374	0.257
	Class 3	0.385	0.422	-	0.446	0.286
	Class 4	0.347	0.374	0.446	-	0.318
	Class 5	0.252	0.257	0.286	0.318	-

Table 5.38: Area under the curve for class versus class performance of the feature ACPeakLocs1.

5.6.2 Location of Second Autocorrelation Peak (ACPeakLocs2)

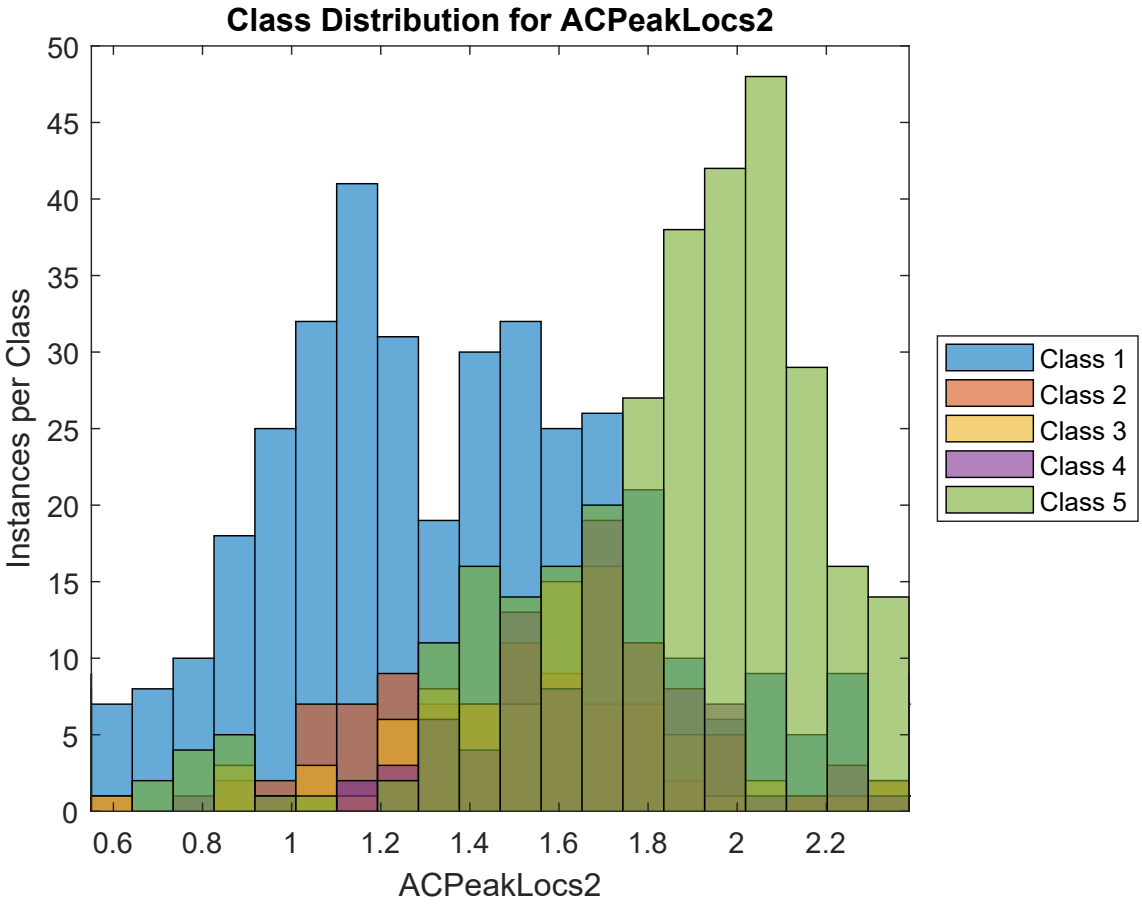

Figure 5.43: Histogram of the class distribution of the ACPeakLocs2 feature.

The ACPeakLocs2 feature is the location of the second peak in the correlogram for each data segment. The location is represented as the time lag at which the peak occurs. A histogram depicting the class distribution of the feature is shown in Figure 5.43. According to the histogram, the feature does not provide any noticeable discriminability between classes 2,3, and 4. The feature does provide discriminability between classes 1 and 5, to an extent. This is corroborated by the table of AUC values in Table 5.39, in which the AUC for class 1 vs class 5 is 0.177. Furthermore,

the table indicates that there is moderate separation between class 5 from classes 2 and 3.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.398	0.325	0.255	0.177
	Class 2	0.398	-	0.404	0.319	0.208
	Class 3	0.325	0.404	-	0.406	0.250
	Class 4	0.255	0.319	0.406	-	0.314
	Class 5	0.177	0.208	0.250	0.314	-

Table 5.39: Area under the curve for class versus class performance of the feature ACPeakLocs2.

5.6.3 Location of Third Autocorrelation Peak (ACPeakLocs3)

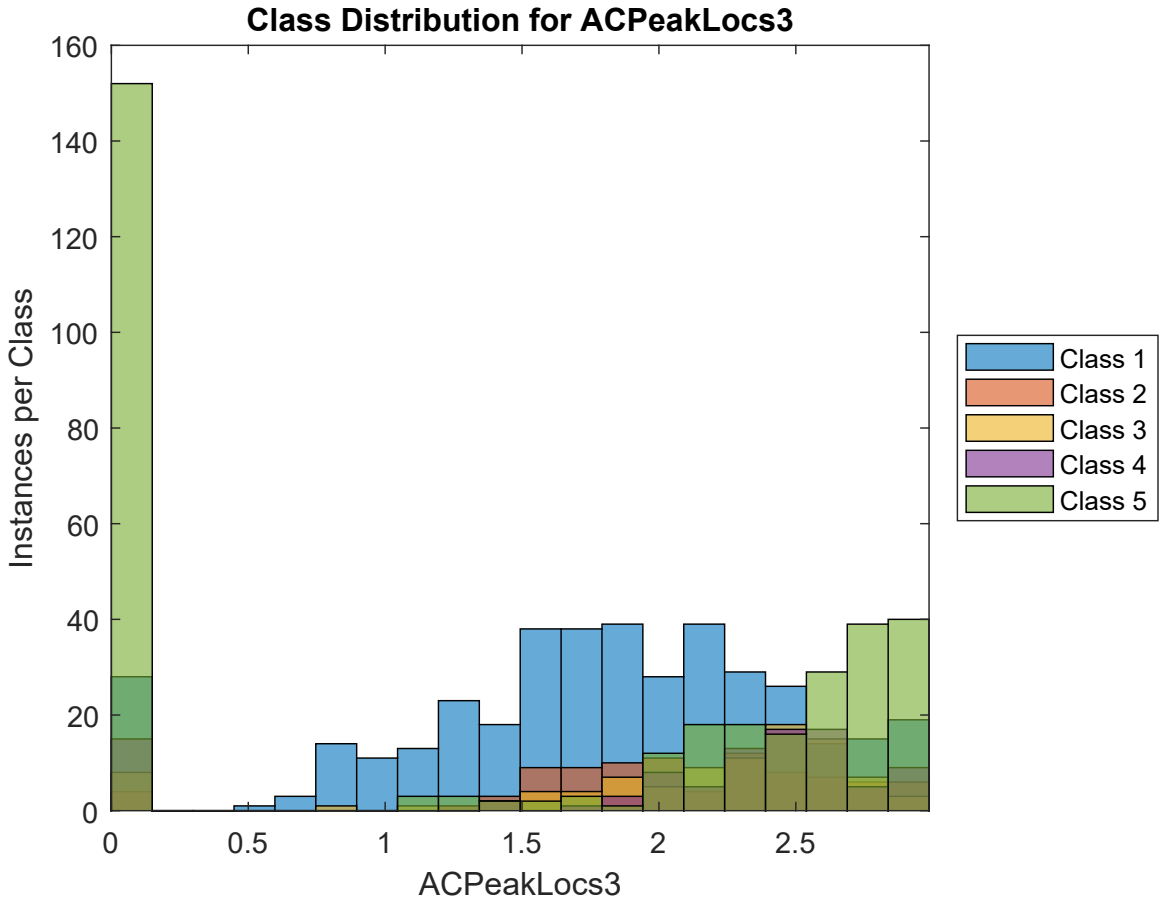

Figure 5.44: Histogram of the class distribution of the ACPeakLocs3 feature.

The ACPeakLocs3 feature is the location of the third peak in the correlogram for each data segment. The location is represented as the time lag at which the peak occurs. A histogram depicting the class distribution of the feature is shown in Figure 5.44. According to the histogram, the feature does not provide any noticeable discriminability between any of the classes. This is corroborated by the table of AUC values in Table 5.40, in which high AUC values are seen for all class combinations.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.369	0.327	0.316	0.470
	Class 2	0.369	-	0.437	0.409	0.419
	Class 3	0.327	0.437	-	0.464	0.408
	Class 4	0.316	0.409	0.464	-	0.403
	Class 5	0.470	0.419	0.408	0.403	-

Table 5.40: Area under the curve for class versus class performance of the feature ACPeakLocs3.

5.6.4 Value of First Autocorrelation Peak (ACPeakVals1)

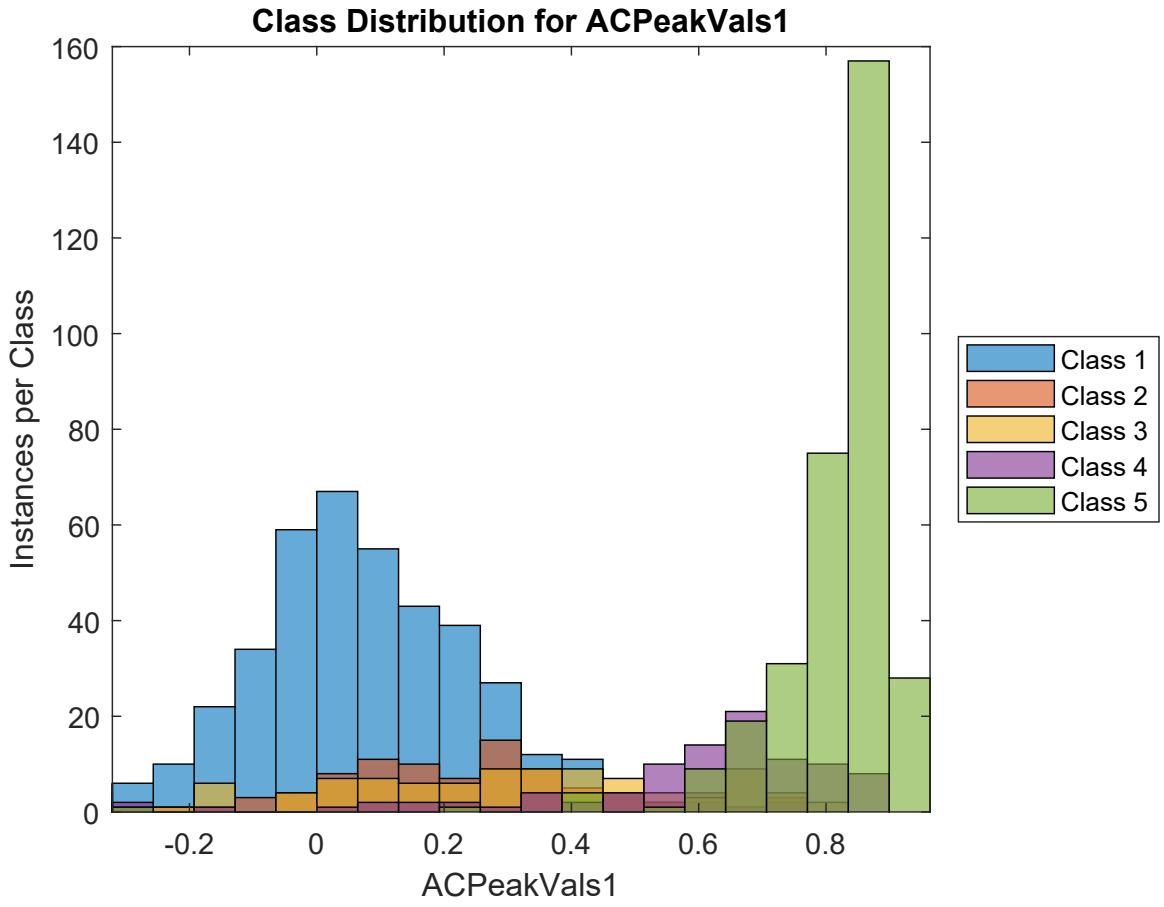

Figure 5.45: Histogram of the class distribution of the ACPeakVals1 feature.

The ACPeakVals1 feature is the autocorrelation value at the first peak in the

correlogram for each data segment. A histogram depicting the class distribution of the feature is shown in Figure 5.45. According to the histogram, the feature provides high discriminability between classes 1 and 5. This is similar to results from our preliminary study [2]. The table of AUC values in Table 5.41 further indicates high discriminability of classes 1, 4, and 5 from each other, as well as from all other classes. The poorest performance is seen between classes 2 and 3.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.286	0.254	0.067	0.040
	Class 2	0.286	-	0.416	0.130	0.046
	Class 3	0.254	0.416	-	0.201	0.059
	Class 4	0.067	0.130	0.201	-	0.154
	Class 5	0.040	0.046	0.059	0.154	-

Table 5.41: Area under the curve for class versus class performance of the feature ACPeakVals1.

5.6.5 Value of Second Autocorrelation Peak (ACPeakVals2)

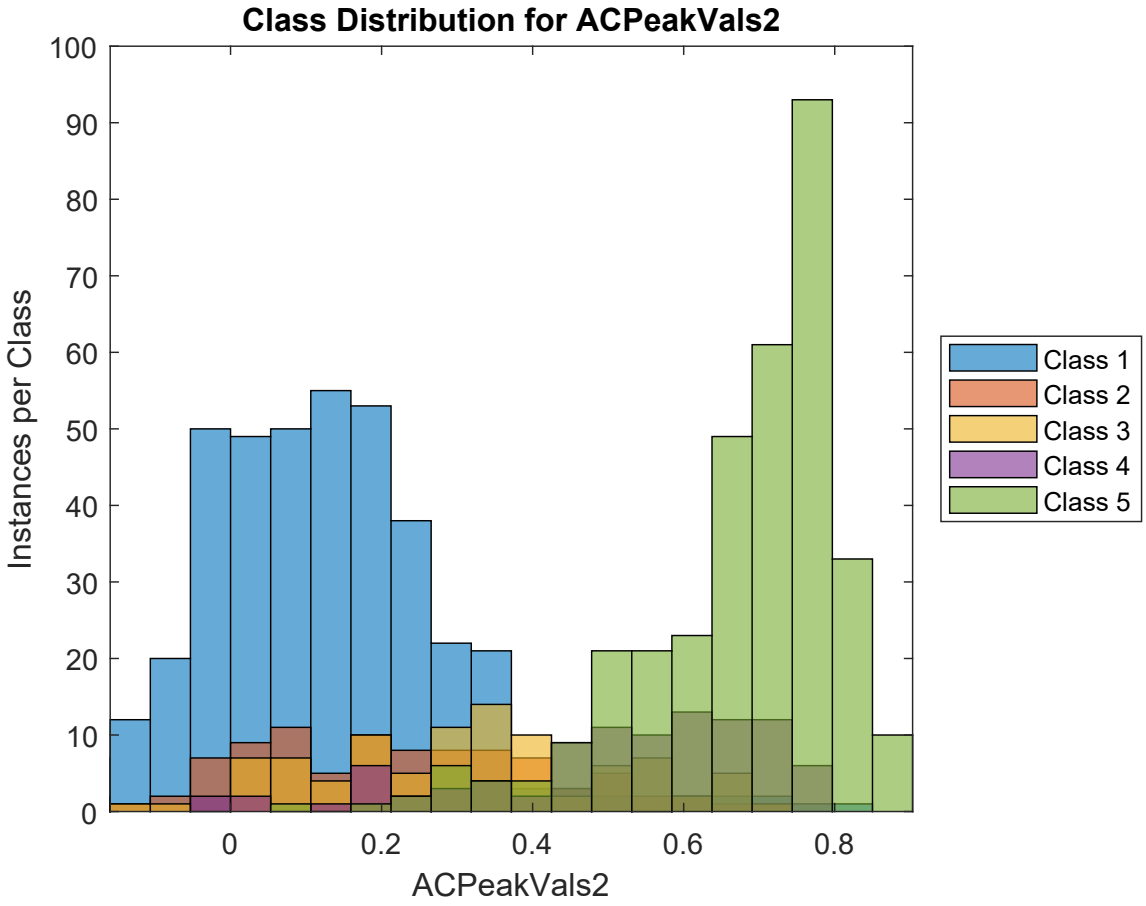

Figure 5.46: Histogram of the class distribution of the ACPeakVals2 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.334	0.220	0.078	0.015
	Class 2	0.334	-	0.379	0.151	0.035
	Class 3	0.220	0.379	-	0.221	0.063
	Class 4	0.078	0.151	0.221	-	0.225
	Class 5	0.015	0.035	0.063	0.225	-

Table 5.42: Area under the curve for class versus class performance of the feature ACPeakVals2.

The ACPeakVals2 feature is the autocorrelation value at the second peak in the correlogram for each data segment. A histogram depicting the class distribution of the feature is shown in Figure 5.46. According to the histogram, the feature provides high discriminability between classes 1 and 5. This is similar to results from our preliminary study [2]. The table of AUC values in Table 5.42 indicates results similar to the ACPeakVals1 feature, with the notable exception of the poor discrimination between class 1 and 2 combination, in which there is an AUC value of 0.334.

5.6.6 Value of Third Autocorrelation Peak (ACPeakVals3)

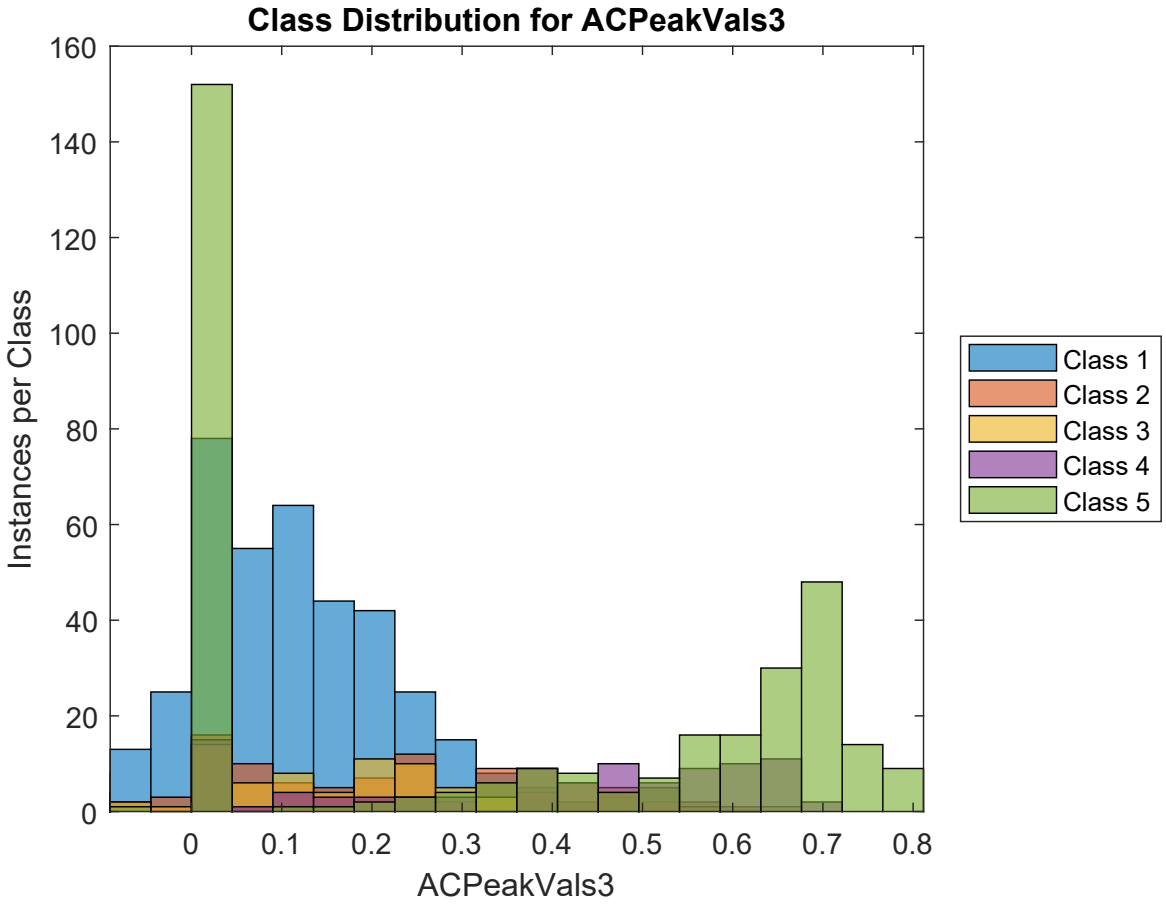

Figure 5.47: Histogram of the class distribution of the ACPeakVals3 feature.

The ACPeakVals3 feature is the autocorrelation value at the third peak in the correlogram for each data segment. A histogram depicting the class distribution of the feature is shown in Figure 5.47. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.43. The only class combination for which there is good discrimination is between classes 1 and 4.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.377	0.332	0.189	0.410
	Class 2	0.377	-	0.454	0.256	0.449
	Class 3	0.332	0.454	-	0.284	0.454
	Class 4	0.189	0.256	0.284	-	0.456
	Class 5	0.410	0.449	0.454	0.456	-

Table 5.43: Area under the curve for class versus class performance of the feature ACPeakVals3.

5.7 Singular Spectrum Analysis Features (Toeplitz Approach)

Functions can be represented as a sum of periodic functions of different periods. Singular spectrum analysis (SSA) is a technique in which the original time series is decomposed into individual components which can be interpreted as trend, oscillatory component, and noise. [34] SSA was performed on each data segment using the Toeplitz and Trajectory approaches, with the resulting eigenvalues used as features. In the Toeplitz approach, a covariance matrix is created by taking the cross-correlation of the signal at various lags. In the trajectory approach, the covariance matrix is created from the scalar product of a time-delayed embedding of the signal. The eigenvalues from the resulting covariance matrices are then used as features. This is implemented in Matlab using the *SSA_beginners_guide_v7* function from the Matlab File Exchange [35].

5.7.1 Toeplitz Eigenvalue 1 (Toep1)

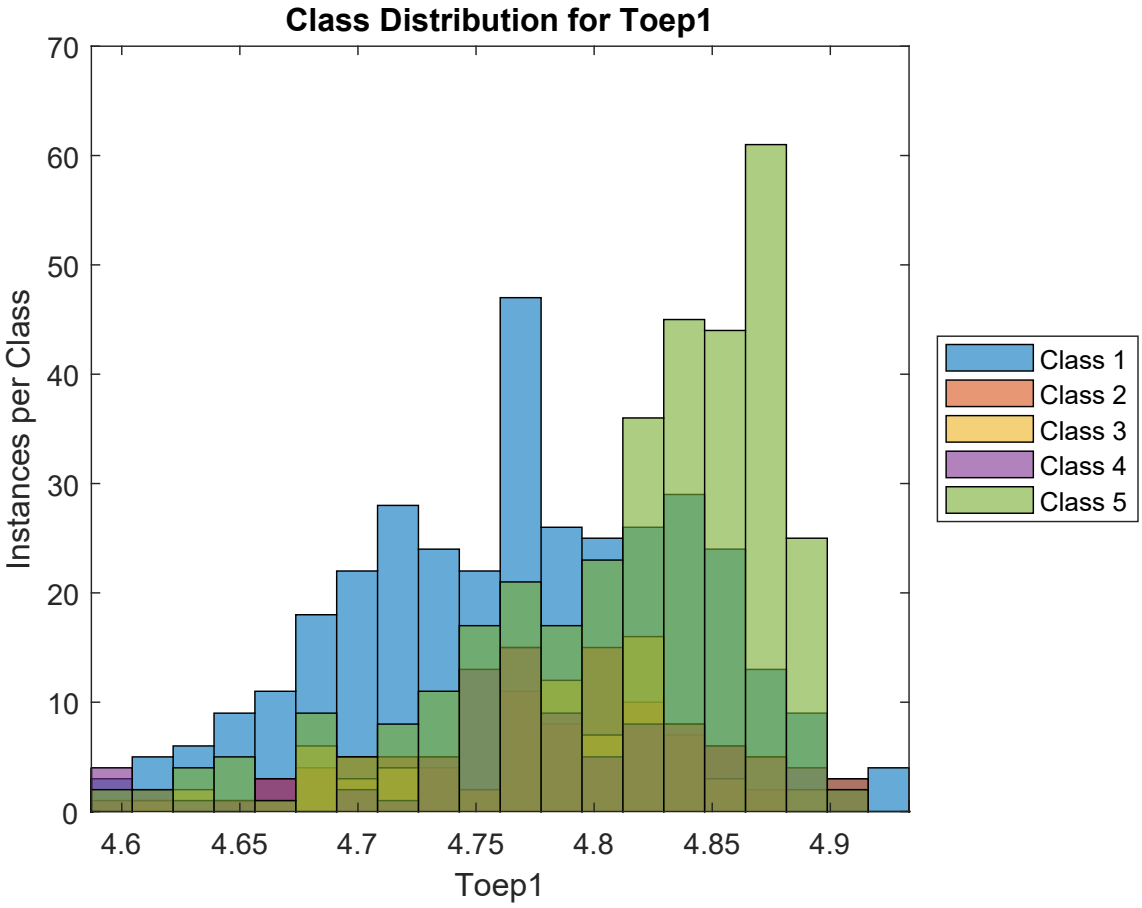

Figure 5.48: Histogram of the class distribution of the Toep1 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.415	0.408	0.412	0.277
	Class 2	0.415	-	0.496	0.497	0.322
	Class 3	0.408	0.496	-	0.491	0.325
	Class 4	0.412	0.497	0.491	-	0.343
	Class 5	0.277	0.322	0.325	0.343	-

Table 5.44: Area under the curve for class versus class performance of the feature Toep1.

The Toep1 feature is the first eigenvalue obtained from the Toeplitz approach. A histogram depicting the class distribution of the feature is shown in Figure 5.48. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.44. The only combination for which there is moderate discrimination is between classes 1 and 5.

5.7.2 Toeplitz Eigenvalue 2 (Toep2)

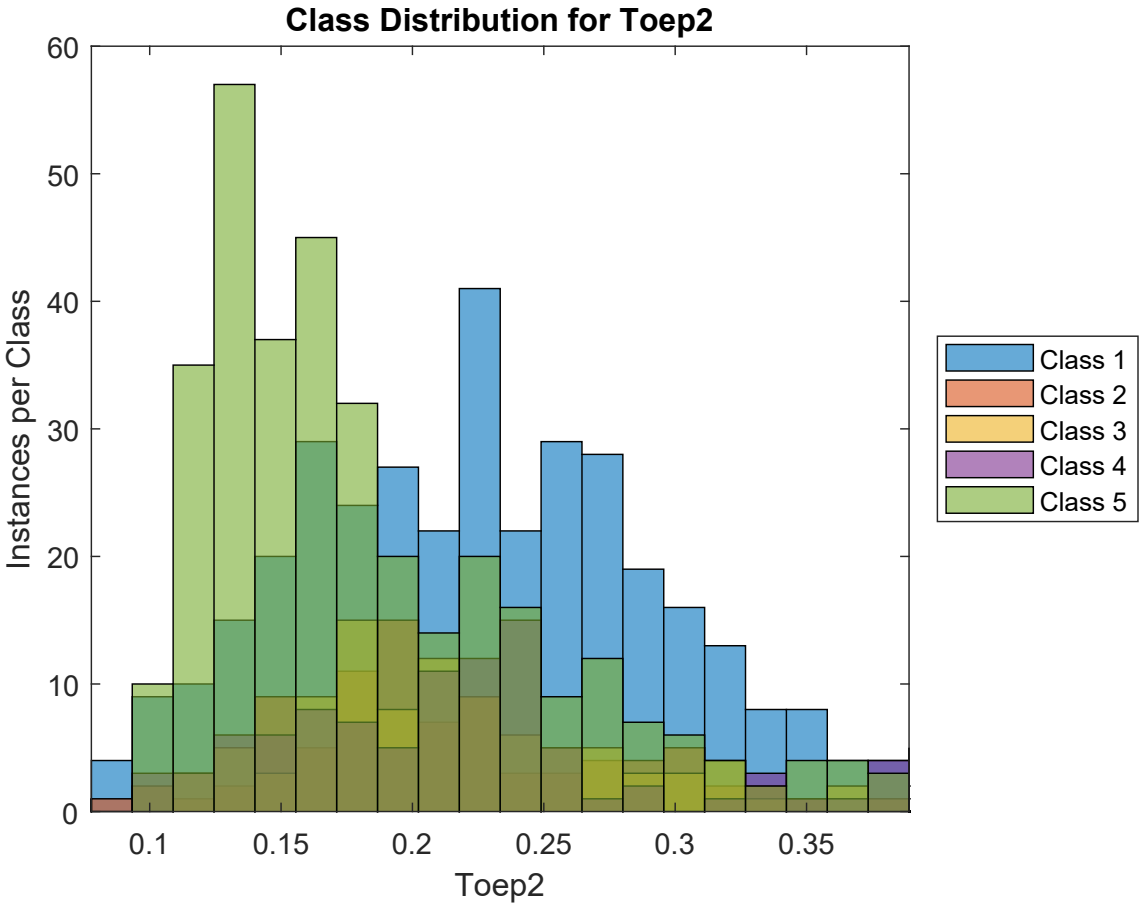

Figure 5.49: Histogram of the class distribution of the Toep2 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.416	0.405	0.406	0.281
	Class 2	0.416	-	0.492	0.497	0.327
	Class 3	0.405	0.492	-	0.493	0.332
	Class 4	0.406	0.497	0.493	-	0.347
	Class 5	0.281	0.327	0.332	0.347	-

Table 5.45: Area under the curve for class versus class performance of the feature Toep2.

The Toep2 feature is the second eigenvalue obtained from the Toeplitz approach. A histogram depicting the class distribution of the feature is shown in Figure 5.49. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.45. The only combination for which there is moderate discrimination is between classes 1 and 5.

5.7.3 Toeplitz Eigenvalue 3 (Toep3)

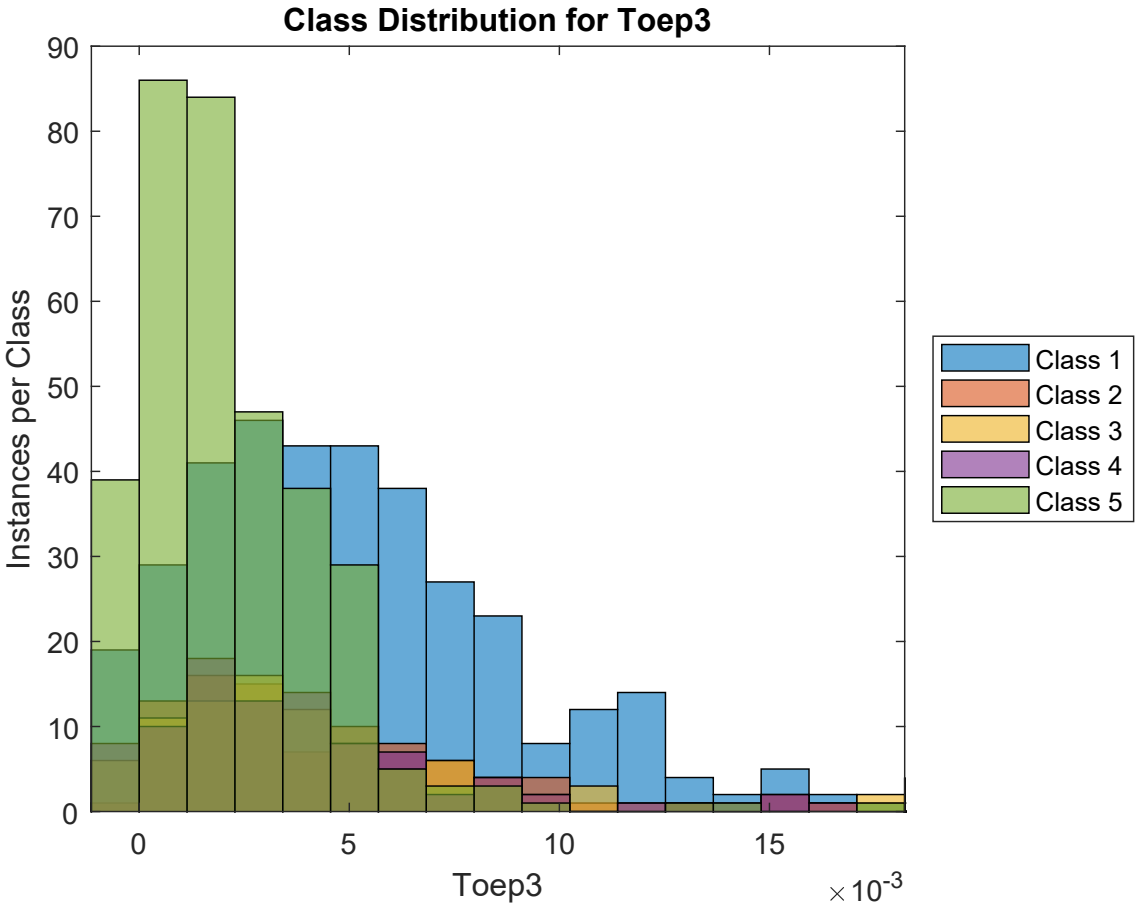

Figure 5.50: Histogram of the class distribution of the Toep3 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.397	0.402	0.367	0.219
	Class 2	0.397	-	0.499	0.465	0.298
	Class 3	0.402	0.499	-	0.465	0.307
	Class 4	0.367	0.465	0.465	-	0.336
	Class 5	0.219	0.298	0.307	0.336	-

Table 5.46: Area under the curve for class versus class performance of the feature Toep3.

The Toep3 feature is the third eigenvalue obtained from the Toeplitz approach. A histogram depicting the class distribution of the feature is shown in Figure 5.50. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.46. The only combination for which there is moderate discrimination is between classes 1 and 5.

5.7.4 Toeplitz Eigenvalue 4 (Toep4)

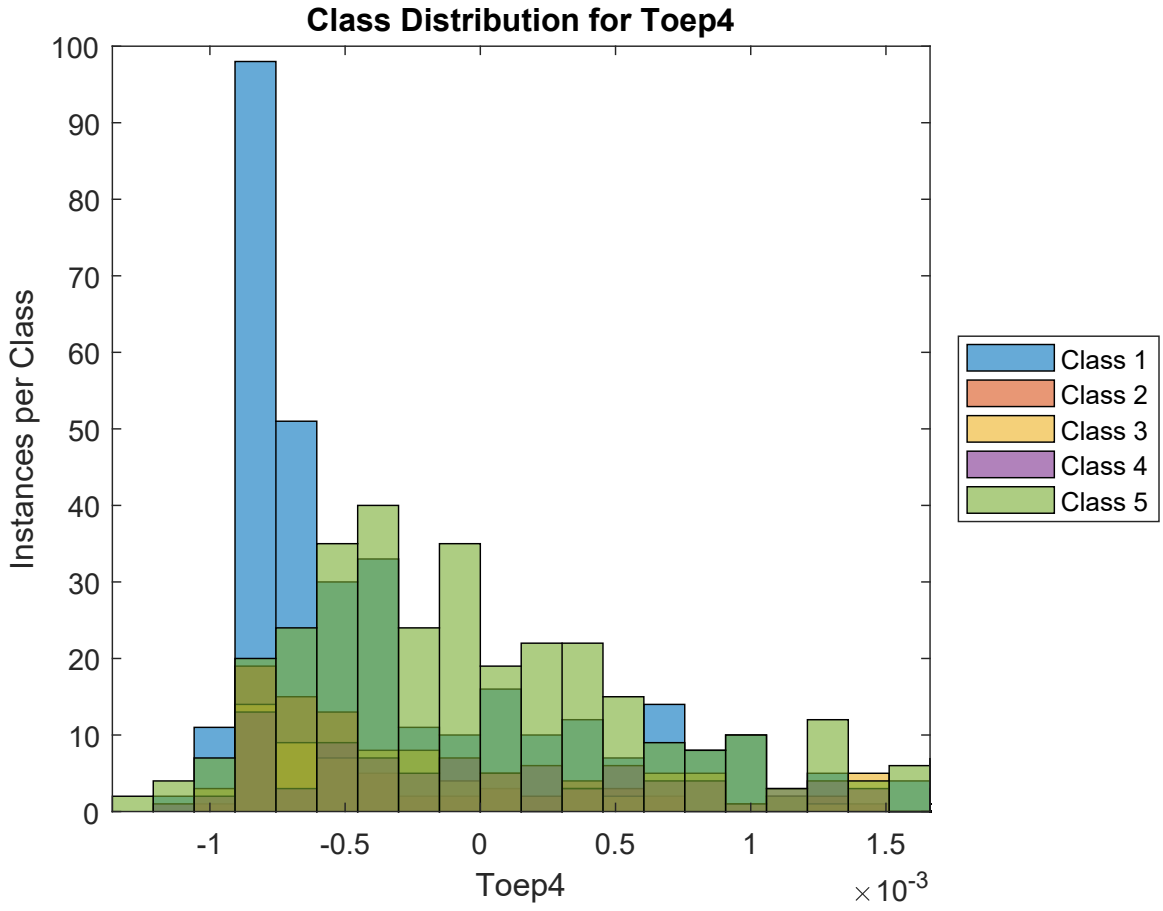

Figure 5.51: Histogram of the class distribution of the Toep4 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.494	0.434	0.415	0.422
	Class 2	0.494	-	0.433	0.417	0.417
	Class 3	0.434	0.433	-	0.481	0.499
	Class 4	0.415	0.417	0.481	-	0.472
	Class 5	0.422	0.417	0.499	0.472	-

Table 5.47: Area under the curve for class versus class performance of the feature Toep4.

The Toep4 feature is the fourth eigenvalue obtained from the Toeplitz approach. A histogram depicting the class distribution of the feature is shown in Figure 5.51. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.47.

5.7.5 Toeplitz Eigenvalue 5 (Toep5)

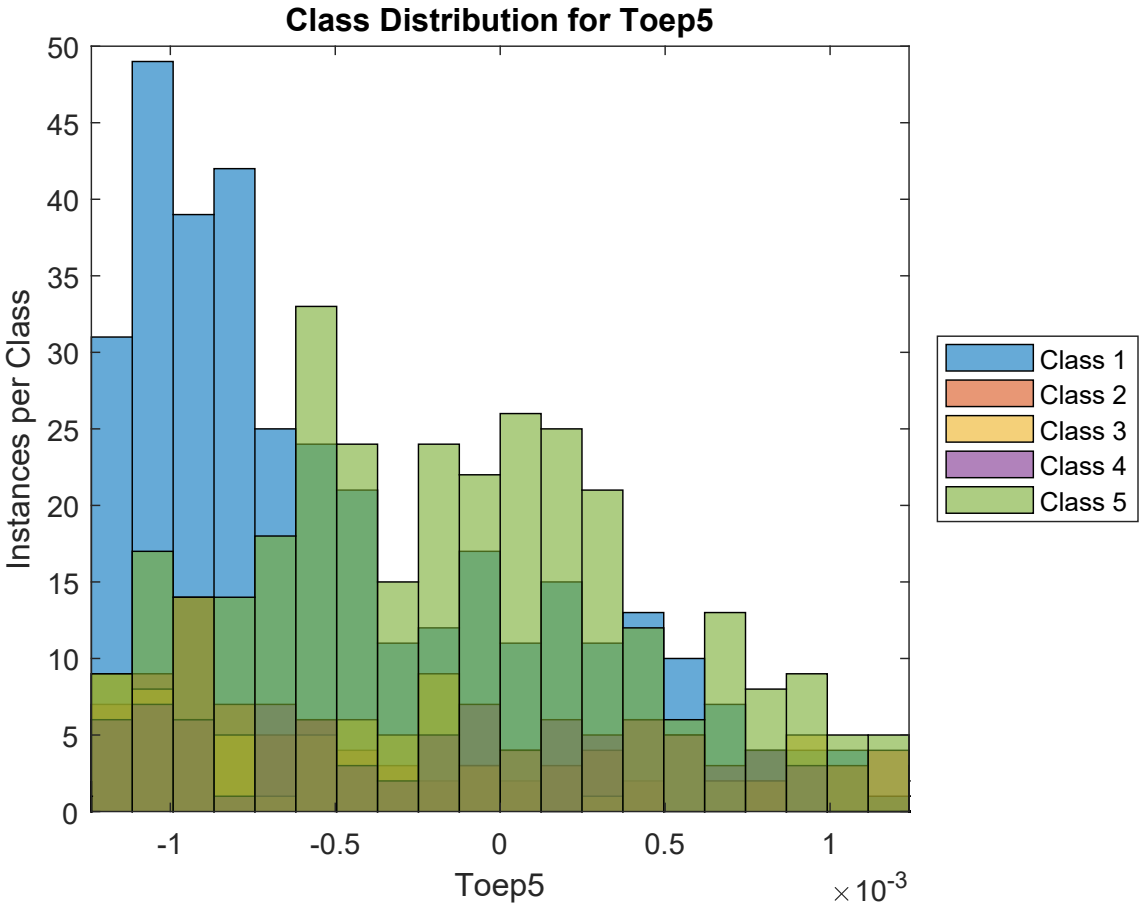

Figure 5.52: Histogram of the class distribution of the Toep5 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.489	0.425	0.397	0.396
	Class 2	0.489	-	0.438	0.416	0.415
	Class 3	0.425	0.438	-	0.480	0.493
	Class 4	0.397	0.416	0.480	-	0.478
	Class 5	0.396	0.415	0.493	0.478	-

Table 5.48: Area under the curve for class versus class performance of the feature Toep5.

The Toep5 feature is the fifth eigenvalue obtained from the Toeplitz approach. A histogram depicting the class distribution of the feature is shown in Figure 5.52. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.48. The only combination for which there is moderate discrimination is between classes 1 and 5.

5.7.6 Trajectory Eigenvalue 1 (Traj1)

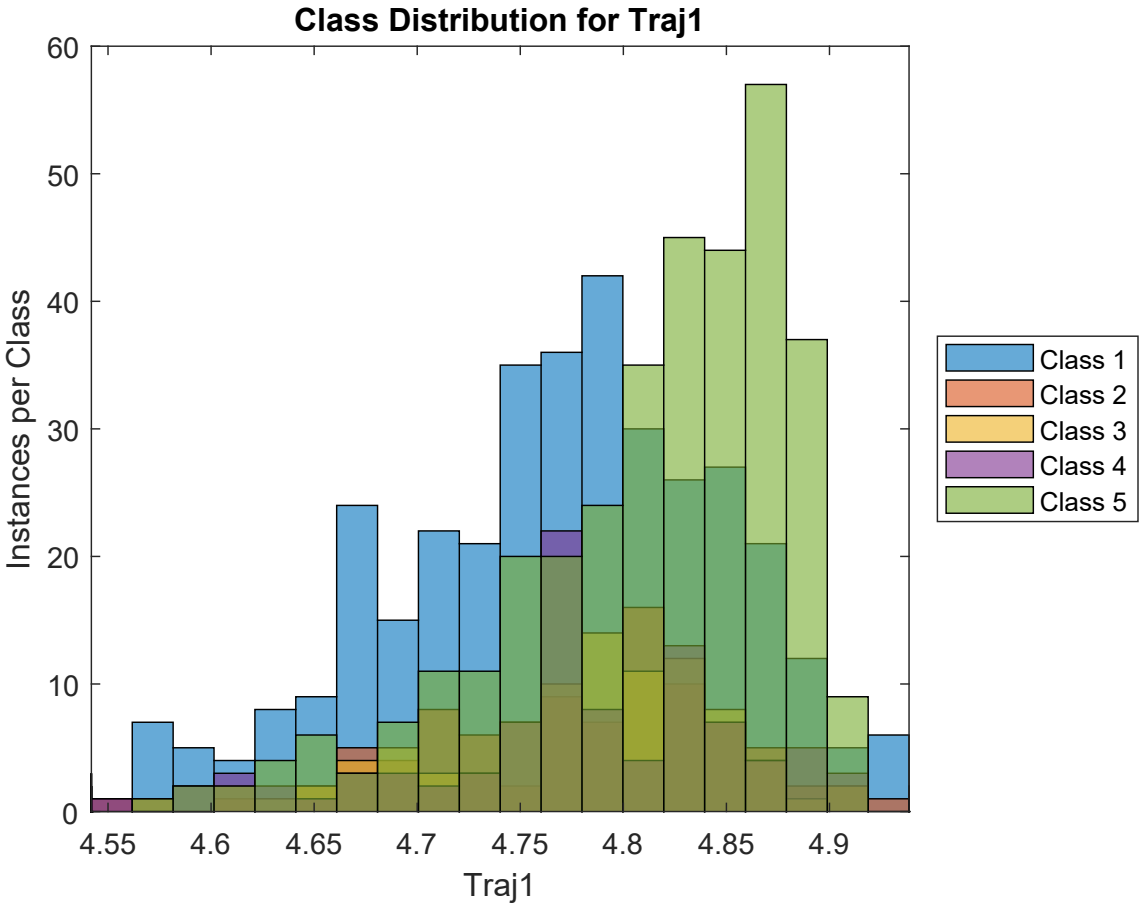

Figure 5.53: Histogram of the class distribution of the Traj1 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.426	0.432	0.430	0.292
	Class 2	0.426	-	0.490	0.496	0.337
	Class 3	0.432	0.490	-	0.499	0.329
	Class 4	0.430	0.496	0.499	-	0.345
	Class 5	0.292	0.337	0.329	0.345	-

Table 5.49: Area under the curve for class versus class performance of the feature Traj1.

The Traj1 feature is the first eigenvalue obtained from the Trajectory approach. A histogram depicting the class distribution of the feature is shown in Figure 5.53. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.49. The only combination for which there is moderate discrimination is between classes 1 and 5.

5.7.7 Trajectory Eigenvalue 2 (Traj2)

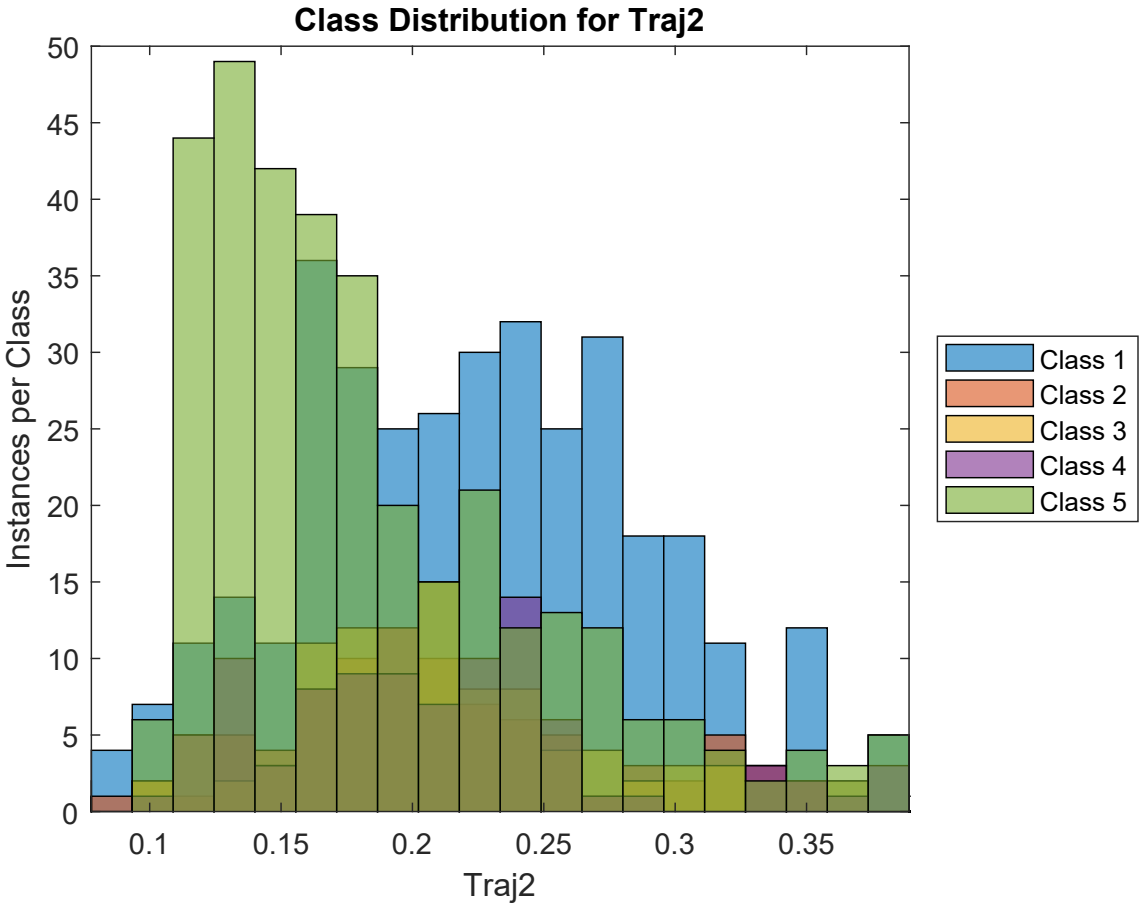

Figure 5.54: Histogram of the class distribution of the Traj2 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.415	0.384	0.394	0.282
	Class 2	0.415	-	0.472	0.486	0.333
	Class 3	0.384	0.472	-	0.487	0.353
	Class 4	0.394	0.486	0.487	-	0.359
	Class 5	0.282	0.333	0.353	0.359	-

Table 5.50: Area under the curve for class versus class performance of the feature Traj2.

The Traj2 feature is the second eigenvalue obtained from the Trajectory approach. A histogram depicting the class distribution of the feature is shown in Figure 5.54. According to the histogram, the feature provides poor discriminability between classes. This is corroborated by the table of AUC values in Table 5.50. The only combination for which there is moderate discrimination is between classes 1 and 5.

5.7.8 Trajectory Eigenvalue 3 (Traj3)

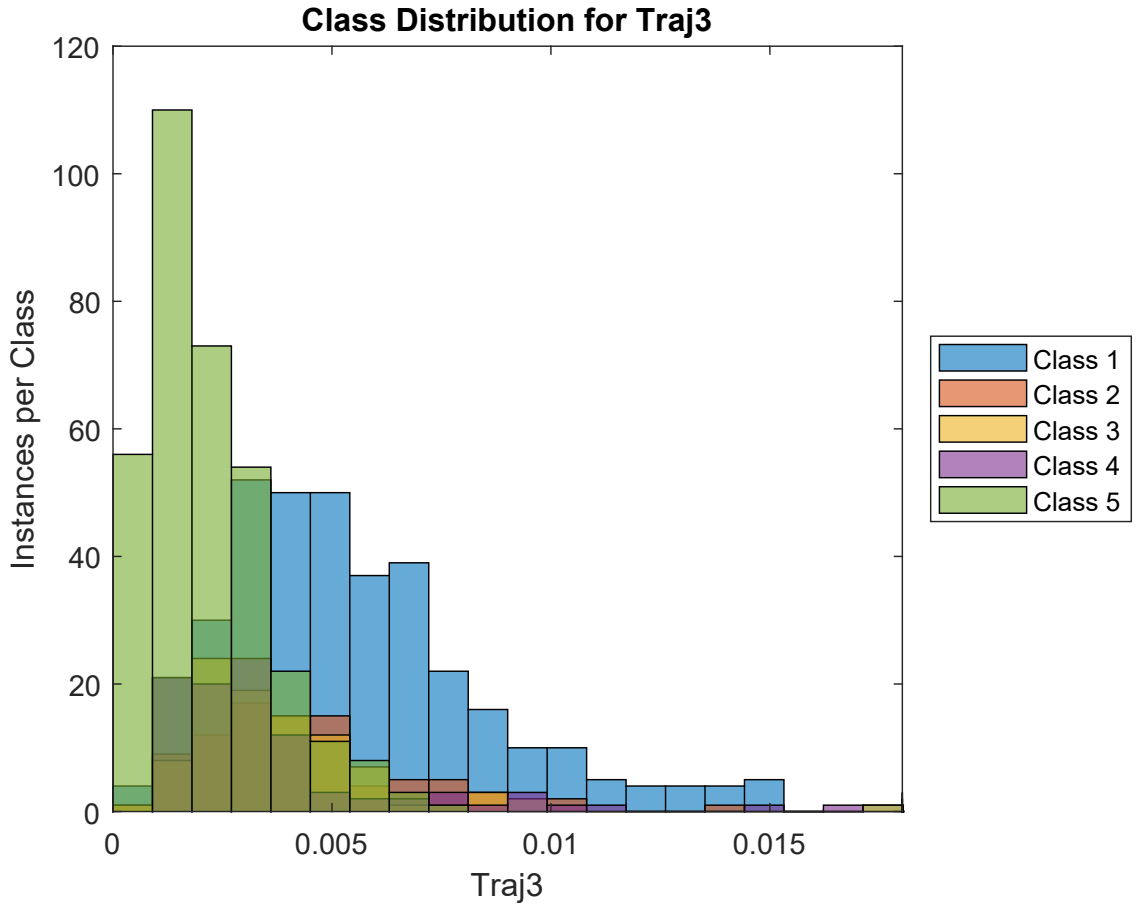

Figure 5.55: Histogram of the class distribution of the Traj3 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.352	0.299	0.257	0.121
	Class 2	0.352	-	0.438	0.365	0.193
	Class 3	0.299	0.438	-	0.407	0.219
	Class 4	0.257	0.365	0.407	-	0.294
	Class 5	0.121	0.193	0.219	0.294	-

Table 5.51: Area under the curve for class versus class performance of the feature Traj3.

The Traj3 feature is the third eigenvalue obtained from the Trajectory approach. A histogram depicting the class distribution of the feature is shown in Figure 5.55. According to the histogram, the feature provides poor discriminability between classes. According to the table of AUC values in Table 5.51, the feature offers good discrimination of class 5 from classes 1 and 2. However, other class combinations yield poor or moderate class discrimination.

5.7.9 Trajectory Eigenvalue 4 (Traj4)

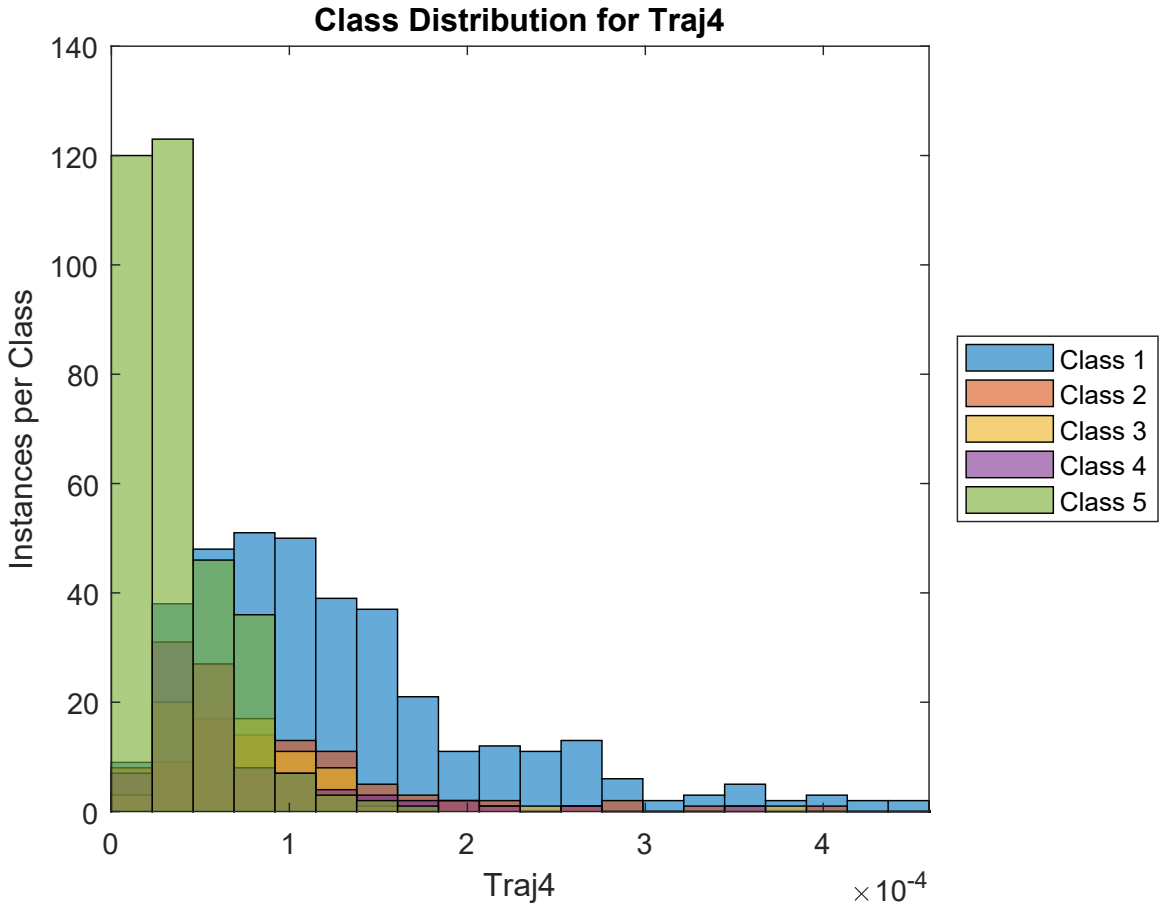

Figure 5.56: Histogram of the class distribution of the Traj4 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.371	0.293	0.242	0.101
	Class 2	0.371	-	0.409	0.343	0.164
	Class 3	0.293	0.409	-	0.408	0.196
	Class 4	0.242	0.343	0.408	-	0.268
	Class 5	0.101	0.164	0.196	0.268	-

Table 5.52: Area under the curve for class versus class performance of the feature Traj4.

The Traj4 feature is the fourth eigenvalue obtained from the Trajectory approach. A histogram depicting the class distribution of the feature is shown in Figure 5.56. According to the histogram, the feature provides poor discriminability between classes. According to the table of AUC values in Table 5.52, the feature offers good discrimination of class 5 from classes 1, 2 and 3. However, other class combinations yield poor or moderate class discrimination.

5.7.10 Trajectory Eigenvalue 5 (Traj5)

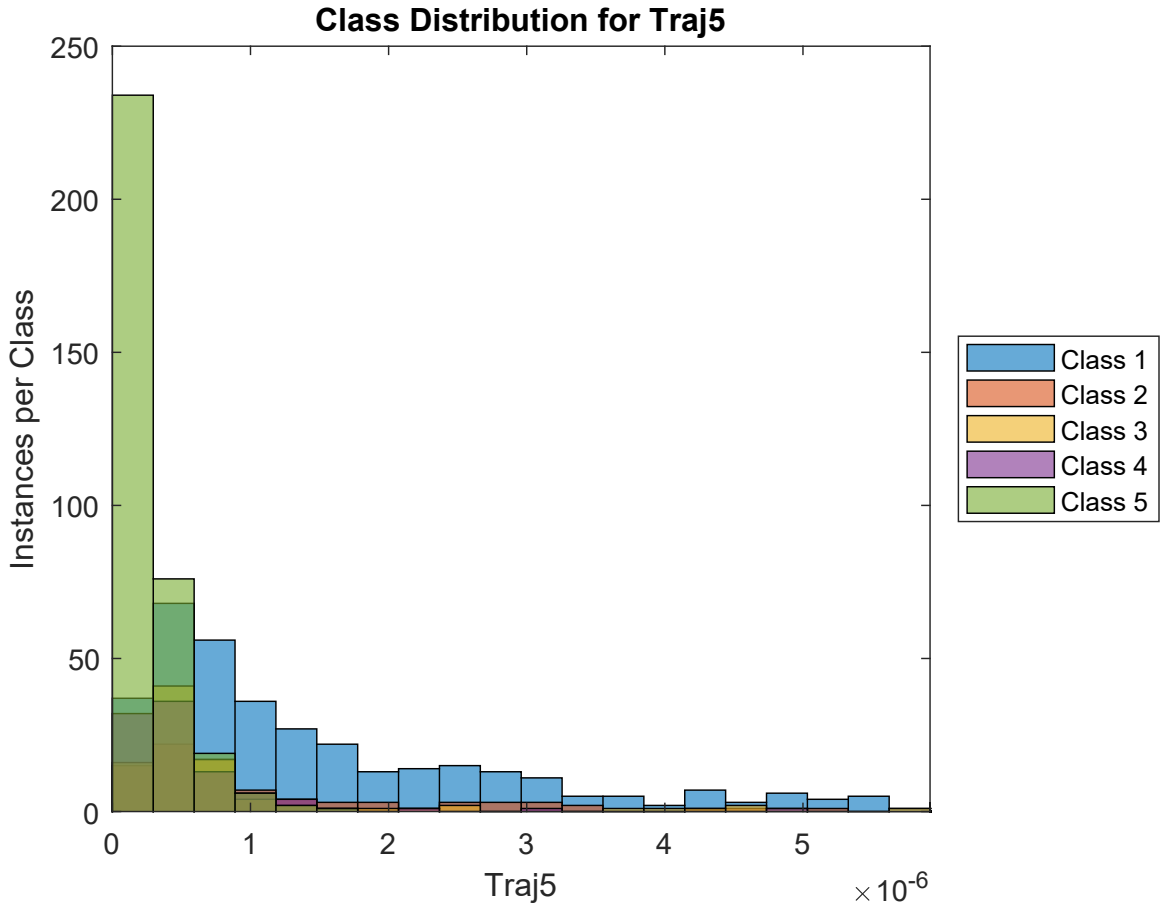

Figure 5.57: Histogram of the class distribution of the Traj5 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.376	0.297	0.206	0.084
	Class 2	0.376	-	0.413	0.313	0.146
	Class 3	0.297	0.413	-	0.388	0.168
	Class 4	0.206	0.313	0.388	-	0.243
	Class 5	0.084	0.146	0.168	0.243	-

Table 5.53: Area under the curve for class versus class performance of the feature Traj5.

The Traj5 feature is the fifth eigenvalue obtained from the Trajectory approach. A histogram depicting the class distribution of the feature is shown in Figure 5.57. According to the histogram, the feature provides poor discriminability between classes. According to the table of AUC values in Table 5.53, the feature offers good discrimination of class 5 from classes 1, 2 and 3. However, other class combinations yield poor or moderate class discrimination.

5.8 Autoregression Model Features

Autoregression (AR) models are used for the prediction of stationary time series [36]. As high quality PPG signals are expected to be stationary, the AR coefficients and the fit of the quality of the fit of the AR model to the actual data were used as features. In AR, the future value of the PPG is predicted using a linear combination of its past values [36]. An AR(4) model was fitted to the data segments using the *ar* function in Matlab.

5.8.1 Autoregression Coefficient 1 (AR1)

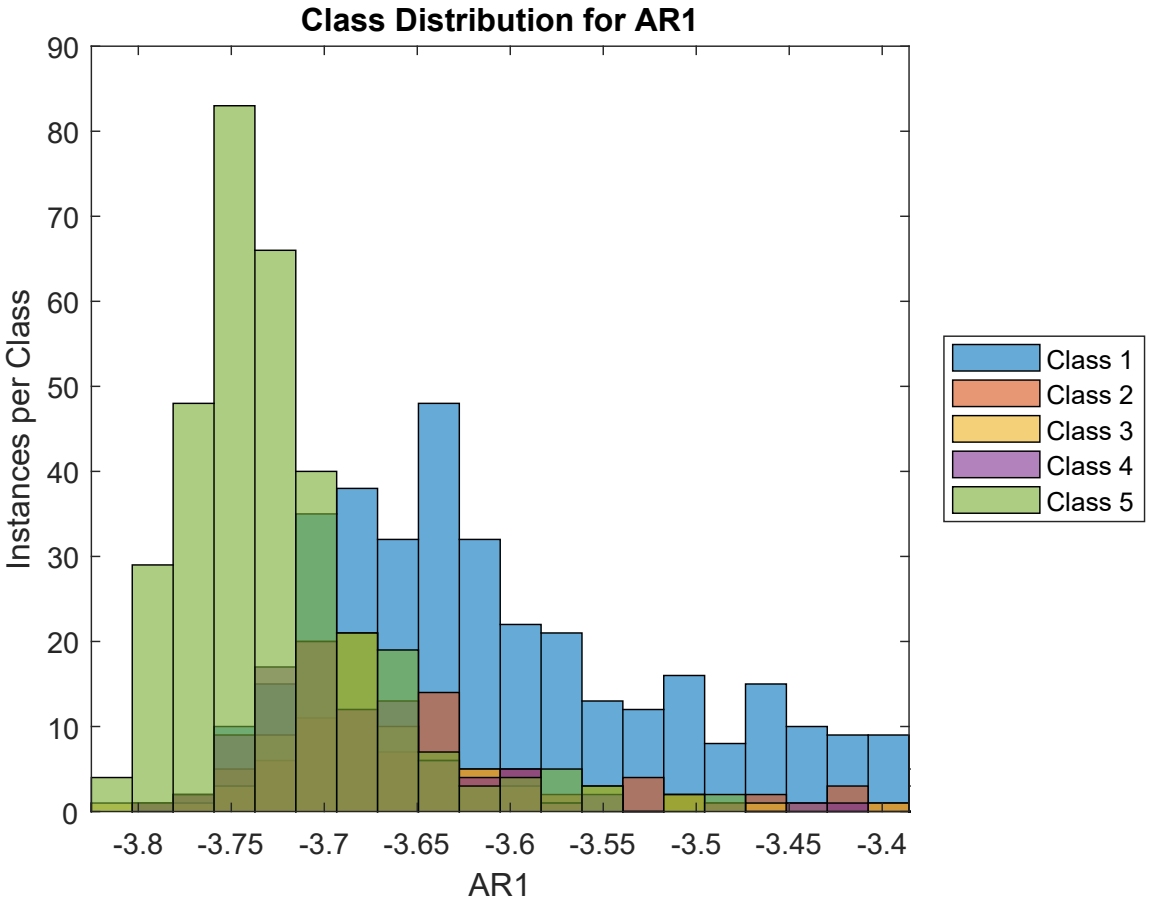

Figure 5.58: Histogram of the class distribution of the AR1 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.386	0.307	0.223	0.101
	Class 2	0.386	-	0.417	0.332	0.164
	Class 3	0.307	0.417	-	0.402	0.190
	Class 4	0.223	0.332	0.402	-	0.257
	Class 5	0.101	0.164	0.190	0.257	-

Table 5.54: Area under the curve for class versus class performance of the feature AR1.

The AR1 feature is the first coefficient of the autoregression model fit. A histogram depicting the class distribution of the feature is shown in Figure 5.58. According to the histogram, the feature provides good discrimination between classes 1 and 5. Additionally, the table of AUC values in Table 5.54 shows that the feature has good discrimination of class 5 from other classes. However the discrimination between other classes is poor.

5.8.2 Autoregression Coefficient 2 (AR2)

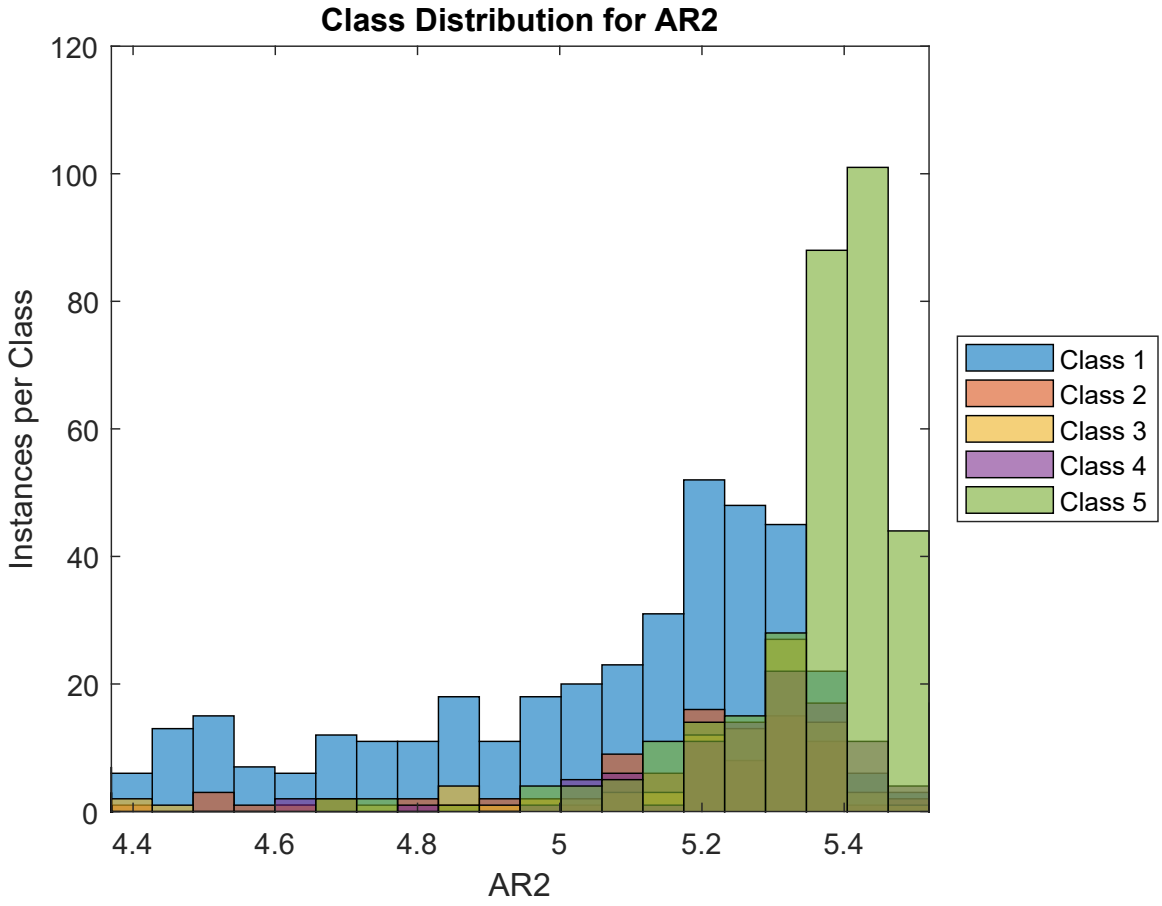

Figure 5.59: Histogram of the class distribution of the AR2 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.394	0.316	0.223	0.107
	Class 2	0.394	-	0.419	0.324	0.166
	Class 3	0.316	0.419	-	0.394	0.193
	Class 4	0.223	0.324	0.394	-	0.262
	Class 5	0.107	0.166	0.193	0.262	-

Table 5.55: Area under the curve for class versus class performance of the feature AR2.

The AR2 feature is the second coefficient of the autoregression model fit. A histogram depicting the class distribution of the feature is shown in Figure 5.59. According to the histogram, the feature provides good discrimination between classes 1 and 5. Additionally, the table of AUC values in Table 5.55 shows that the feature has good discrimination of class 5 from other classes. However the discrimination between other classes is poor.

5.8.3 Autoregression Coefficient 3 (AR3)

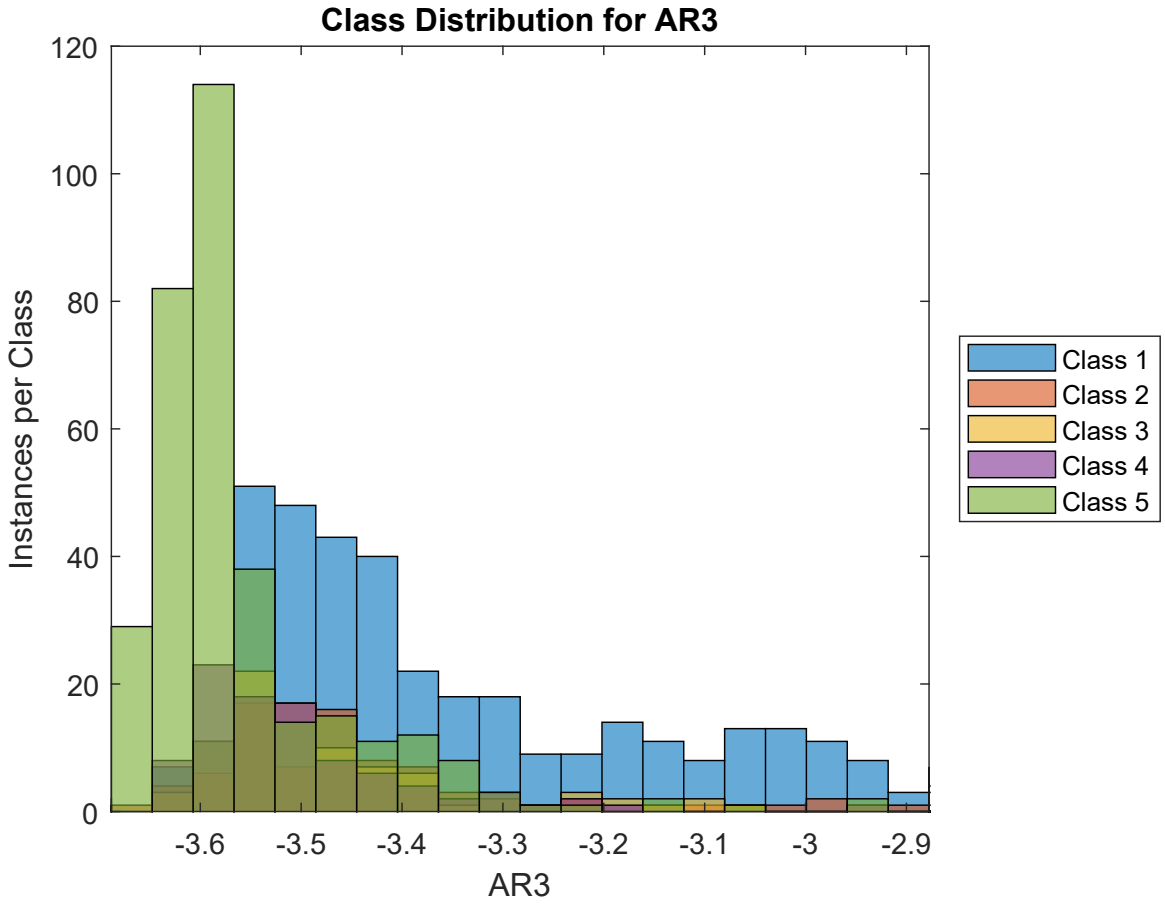

Figure 5.60: Histogram of the class distribution of the AR3 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.406	0.330	0.226	0.122
	Class 2	0.406	-	0.424	0.315	0.181
	Class 3	0.330	0.424	-	0.383	0.211
	Class 4	0.226	0.315	0.383	-	0.292
	Class 5	0.122	0.181	0.211	0.292	-

Table 5.56: Area under the curve for class versus class performance of the feature AR3.

The AR3 feature is the third coefficient of the autoregression model fit. A histogram depicting the class distribution of the feature is shown in Figure 5.60. According to the histogram, the feature provides good discrimination between classes 1 and 5. Additionally, the table of AUC values in Table 5.56 shows that the feature has good discrimination of class 5 from other classes. However the discrimination between other classes is poor.

5.8.4 Autoregression Coefficient 4 (AR4)

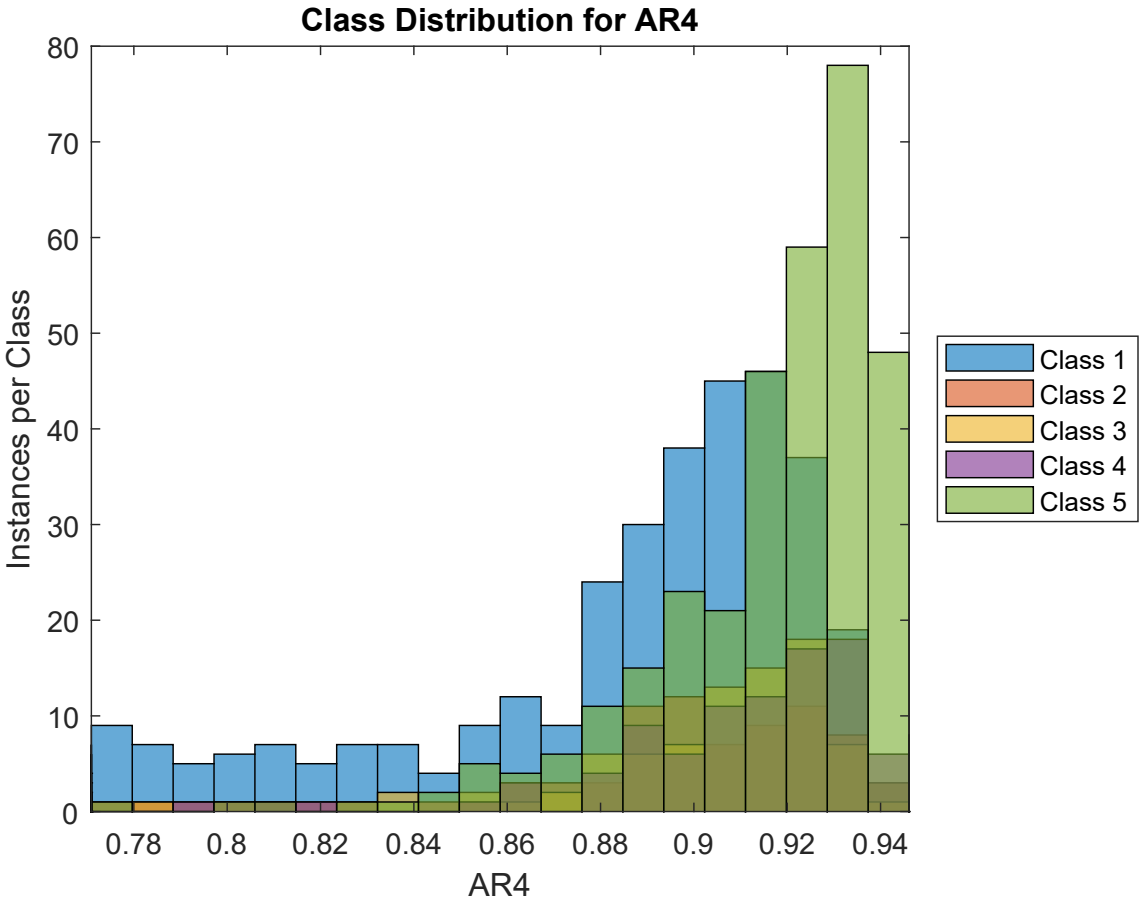

Figure 5.61: Histogram of the class distribution of the AR4 feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.446	0.381	0.269	0.211
	Class 2	0.446	-	0.437	0.324	0.260
	Class 3	0.381	0.437	-	0.376	0.300
	Class 4	0.269	0.324	0.376	-	0.420
	Class 5	0.211	0.260	0.300	0.420	-

Table 5.57: Area under the curve for class versus class performance of the feature AR4.

The AR4 feature is the fourth coefficient of the autoregression model fit. A histogram depicting the class distribution of the feature is shown in Figure 5.61. According to the histogram, the feature does not provide good discrimination between any classes. This is corroborated by the table of AUC values in Table 5.57, which shows poor or moderate discrimination for all class combinations. This feature performs more poorly than the AR1, AR2 and AR3 features.

5.8.5 Autoregression Fit (ARFit)

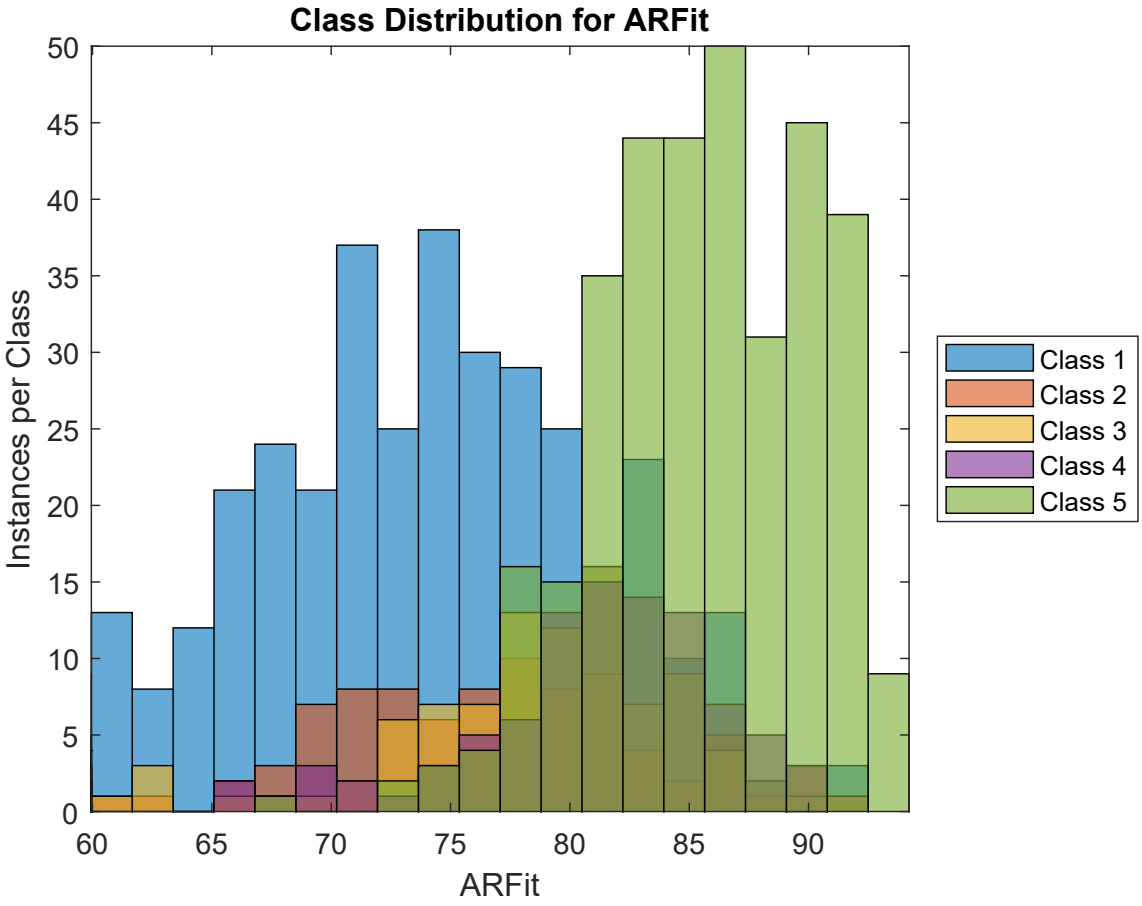

Figure 5.62: Histogram of the class distribution of the ARFit feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality Classes	**Class 1**	-	0.367	0.276	0.205	0.075
	Class 2	0.367	-	0.393	0.300	0.131
	Class 3	0.276	0.393	-	0.377	0.160
	Class 4	0.205	0.300	0.377	-	0.245
	Class 5	0.075	0.131	0.160	0.245	-

Table 5.58: Area Under the curve for class versus class performance of the feature ARFit.

The fit of the AR model to the actual data was used as a feature. It is postulated that the AR model will provide a better fit got higher quality PPG segments. The *compare* function in Matlab was used to compute the effectiveness of the fit, using a prediction 5 steps data points ahead. The fit is computed as a percentage accounting for the difference between the AR model prediction and the actual values, as seen in Equation 5.8.

$$ARFit = 100 \times \left(1 - \frac{\|y - \hat{y}\|}{\|y - mean(y)\|}\right) \tag{5.8}$$

Where $\hat{y}$ are the predicted values and y are the actual values.

A histogram depicting the class distribution of the feature is shown in Figure 5.62. According to the histogram, the feature provides good discrimination between classes 1 and 5. Additionally, the table of AUC values in Table 5.58 shows that the feature has good discrimination of class 5 from other classes. However the discrimination between other classes is poor. Its discrimination between classes 1 and 5 is superior to the other autoregression features, its overall performance for other class combinations is similar.

5.9 Accelerometer Features

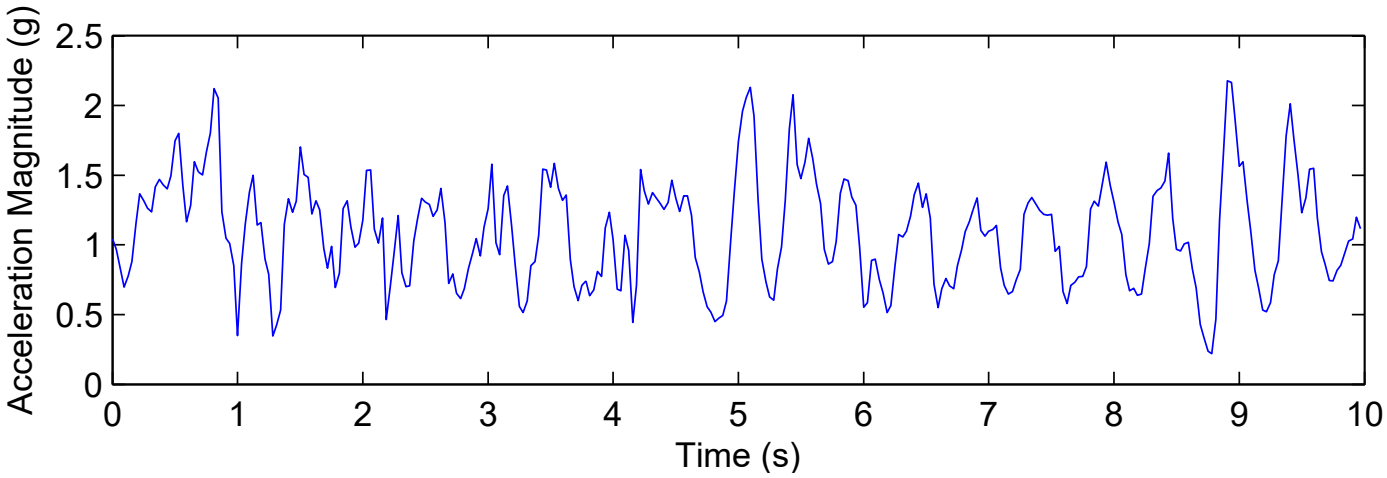

Figure 5.63: Magnitude accelerometer signal for class 1 data segment.

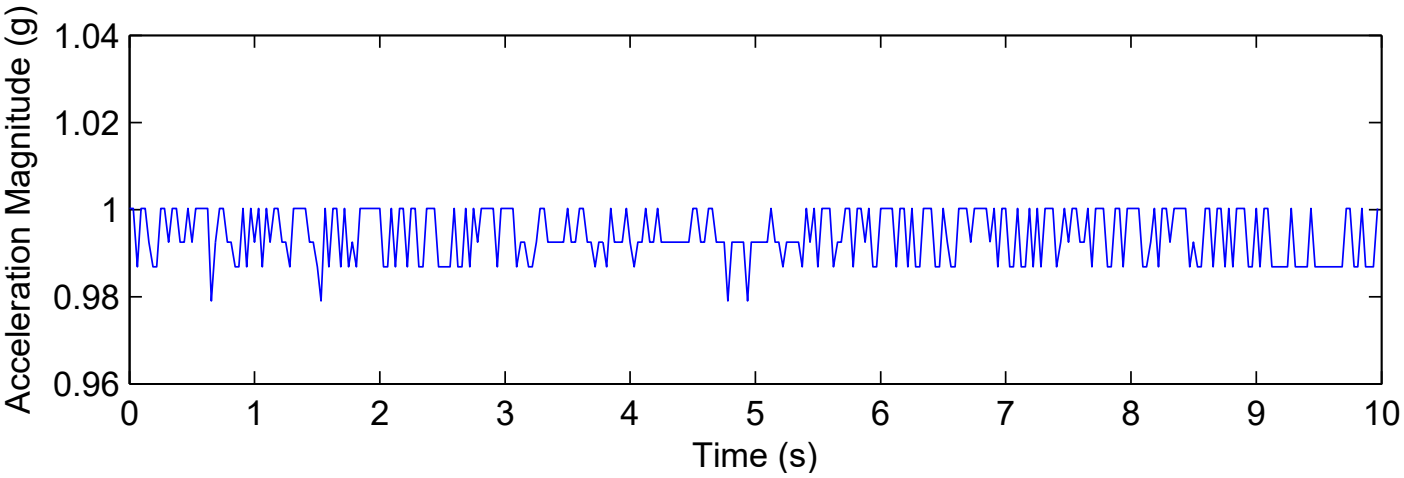

Figure 5.64: Magnitude accelerometer signal for class 5 data segment.

Empatica E4 wristbands contain 3-axis accelerometers that record data at 32Hz, continuously with the PPG sensor. Low quality data is expected to be found at times when the participant engages in any motion of the arm. Such motion would register on the accelerometer in the axes corresponding to the direction of motion. The magnitude of the acceleration is computed using Equation 5.9 to combine the

x, y, and z axes of the accelerometer data. When the magnitude of the acceleration differs from the acceleration due to gravity, it is likely due to movement; hence the accelerometer data ought to correlate with PPG signal quality. This seen in the class 1 accelerometer signal in Figure 5.63, which has a larger range and different shape than the class 5 signal in Figure 5.64.

$$magACC = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{5.9}$$

Where a_x, a_y, and a_z are the accelerations from the x, y, and z axes, respectively.

5.9.1 Mean of Accelerometer Signal Magnitude (meanACC)

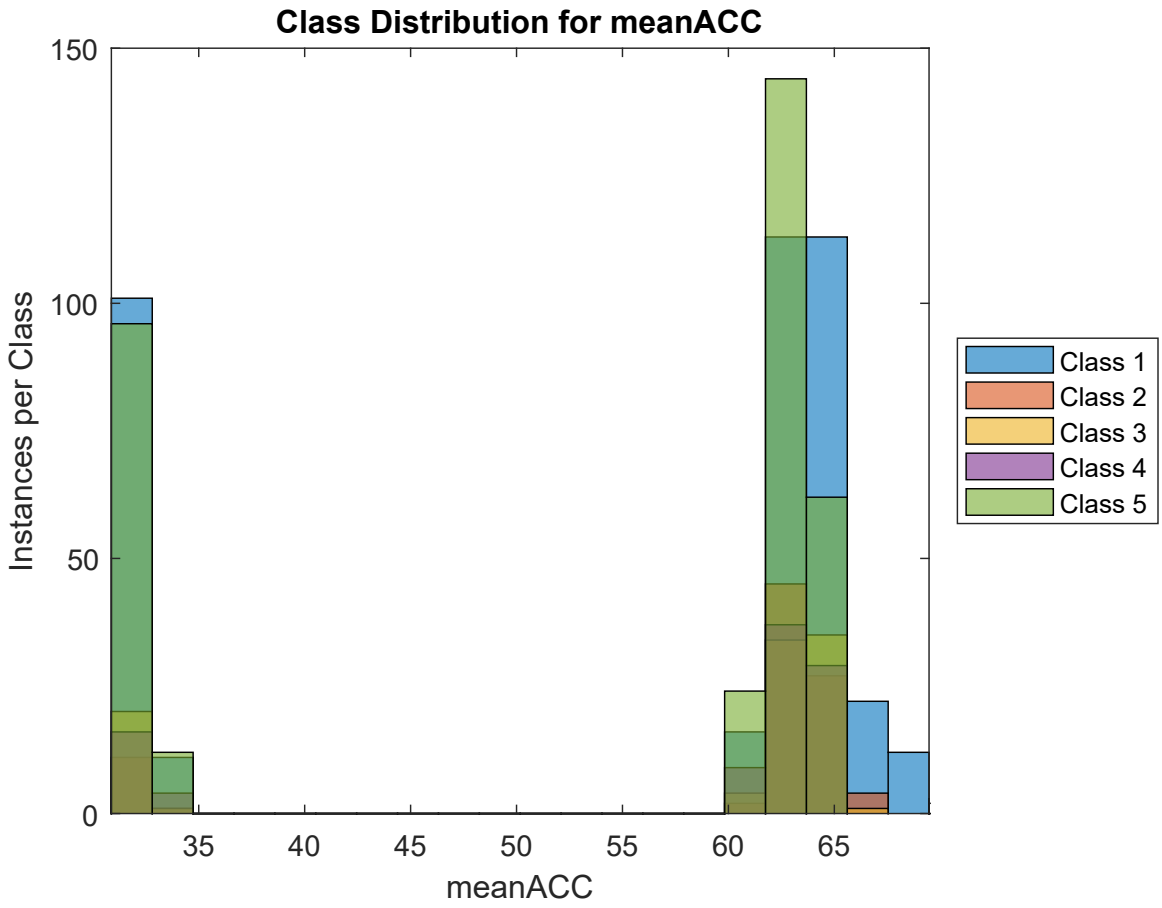

Figure 5.65: Histogram of the class distribution of the meanACC feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.495	0.498	0.472	0.387
	Class 2	0.495	-	0.500	0.454	0.355
	Class 3	0.498	0.500	-	0.463	0.373
	Class 4	0.472	0.454	0.463	-	0.408
	Class 5	0.387	0.355	0.373	0.408	-

Table 5.59: Area under the curve for class versus class performance of the feature meanACC.

The meanACC feature is the mean of the magnitude of the accelerometer signal for the 10 second data segment. A histogram depicting the class distribution of the feature is shown in Figure 5.65. According to the histogram, the feature provides poor class discrimination between classes. This is corroborated by the table of AUC values in Table 5.59.

5.9.2 Median of Accelerometer Signal Magnitude (medianACC)

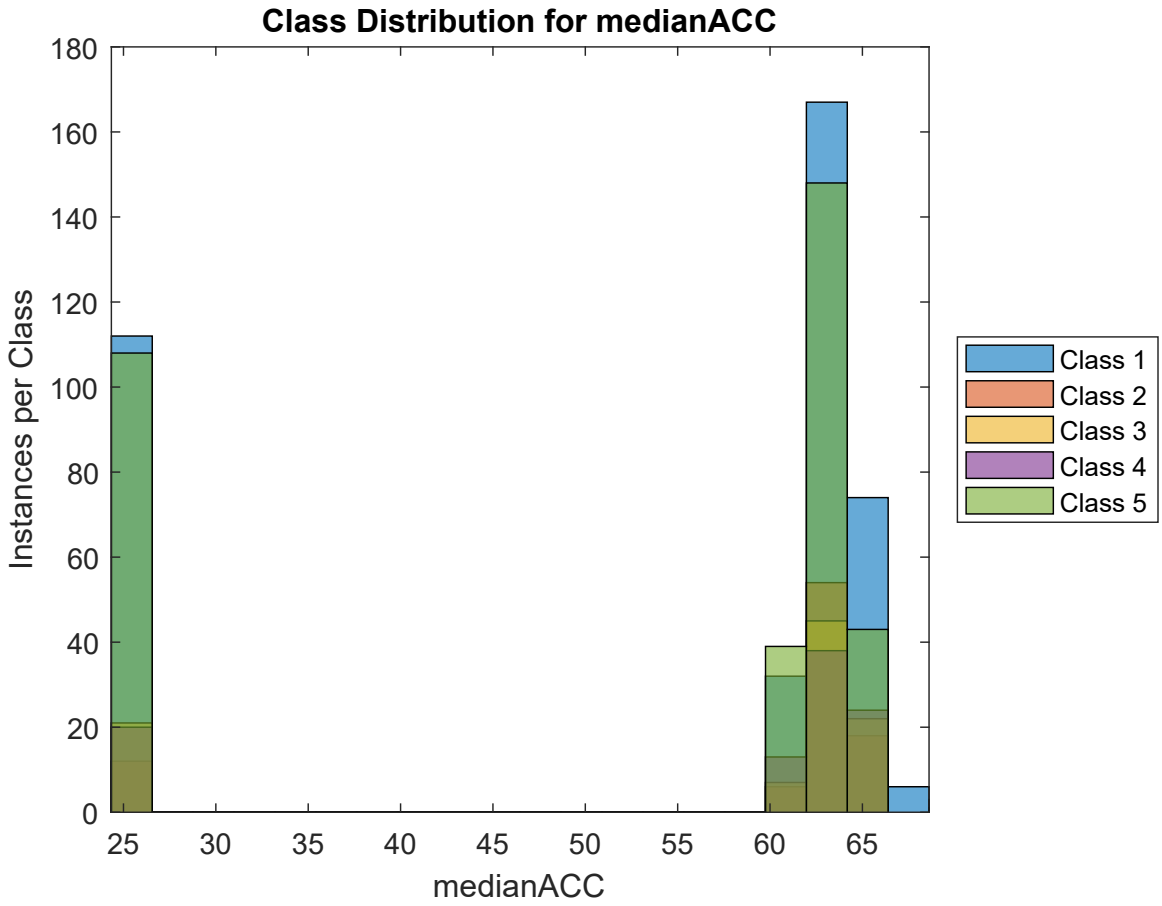

Figure 5.66: Histogram of the class distribution of the medianACC feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.469	0.460	0.483	0.432
	Class 2	0.469	-	0.478	0.475	0.384
	Class 3	0.460	0.478	-	0.473	0.386
	Class 4	0.483	0.475	0.473	-	0.416
	Class 5	0.432	0.384	0.386	0.416	-

Table 5.60: Area under the curve for class versus class performance of the feature medianACC.

The medianACC feature is the median of the magnitude of the accelerometer signal for the 10 second data segment. A histogram depicting the class distribution of the feature is shown in Figure 5.66. According to the histogram, the feature provides poor class discrimination between classes. This is corroborated by the table of AUC values in Table 5.60.

5.9.3 Standard Deviation of Accelerometer Signal Magnitude (stdevACC)

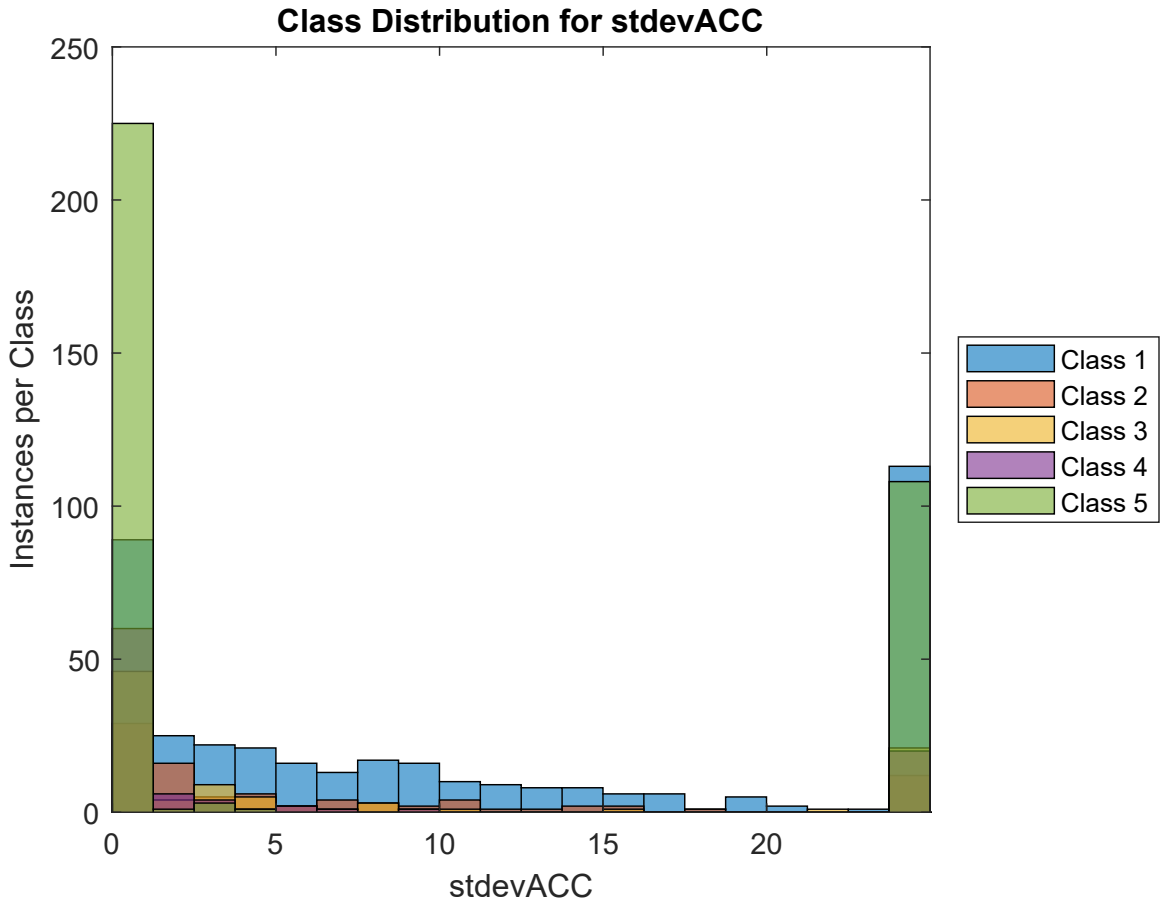

Figure 5.67: Histogram of the class distribution of the stdevACC feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.377	0.366	0.300	0.320
	Class 2	0.377	-	0.453	0.336	0.327
	Class 3	0.366	0.453	-	0.387	0.336
	Class 4	0.300	0.336	0.387	-	0.388
	Class 5	0.320	0.327	0.336	0.388	-

Table 5.61: Area under the curve for class versus class performance of the feature stdevACC.

The stdevACC feature is the standard deviation of the magnitude of the accelerometer signal for the 10 second data segment. A histogram depicting the class distribution of the feature is shown in Figure 5.67. According to the histogram, the feature provides poor class discrimination between classes. This is corroborated by the table of AUC values in Table 5.61.

5.10 Power Spectral Density Features

Differences in periodicity and frequency content are expected between the signal quality classes. Welch periodograms were constructed for each 10 second data segment using the Matlab *pwelch* function to depict the power spectral density (PSD). A window length of 2 seconds was chosen, with an overlap of 1 second between windows. Examples of Welch periodograms constructed for classes 1 and 5 are shown in Figures 5.68 and 5.69, respectively. The power at various frequencies were used as features. These features were inspired by the work of Wander and Morris [21].

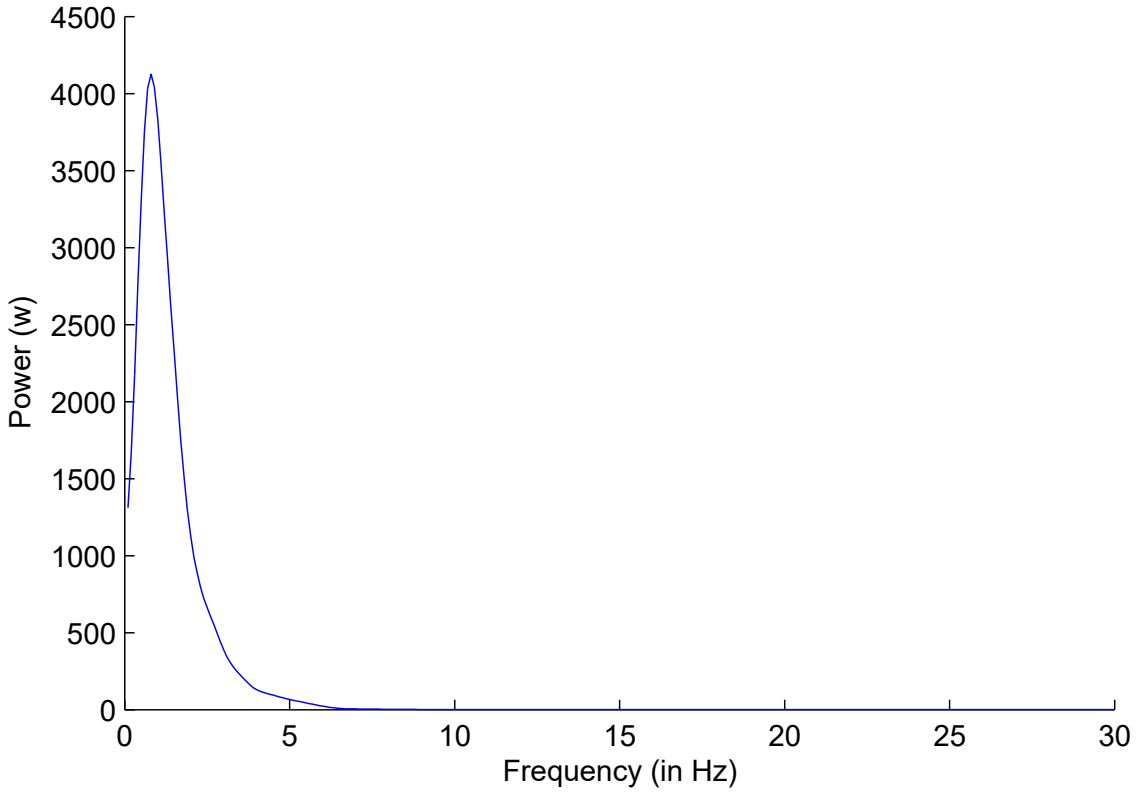

Figure 5.68: Welch periodogram showing the power spectral density for class 1. Reproduced from [2].

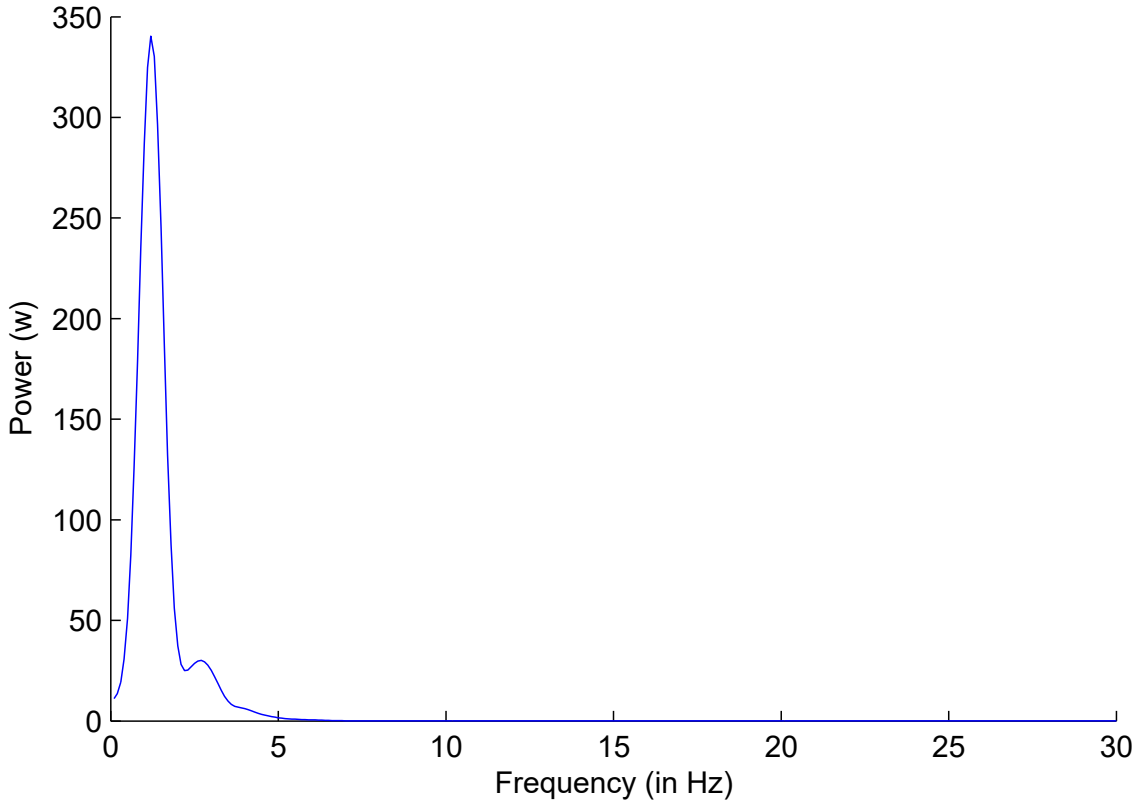

Figure 5.69: Welch periodogram showing the power spectral density for class 5. Reproduced from [2].

5.10.1 Power Spectral Density at 1Hz (PSD1Hz)

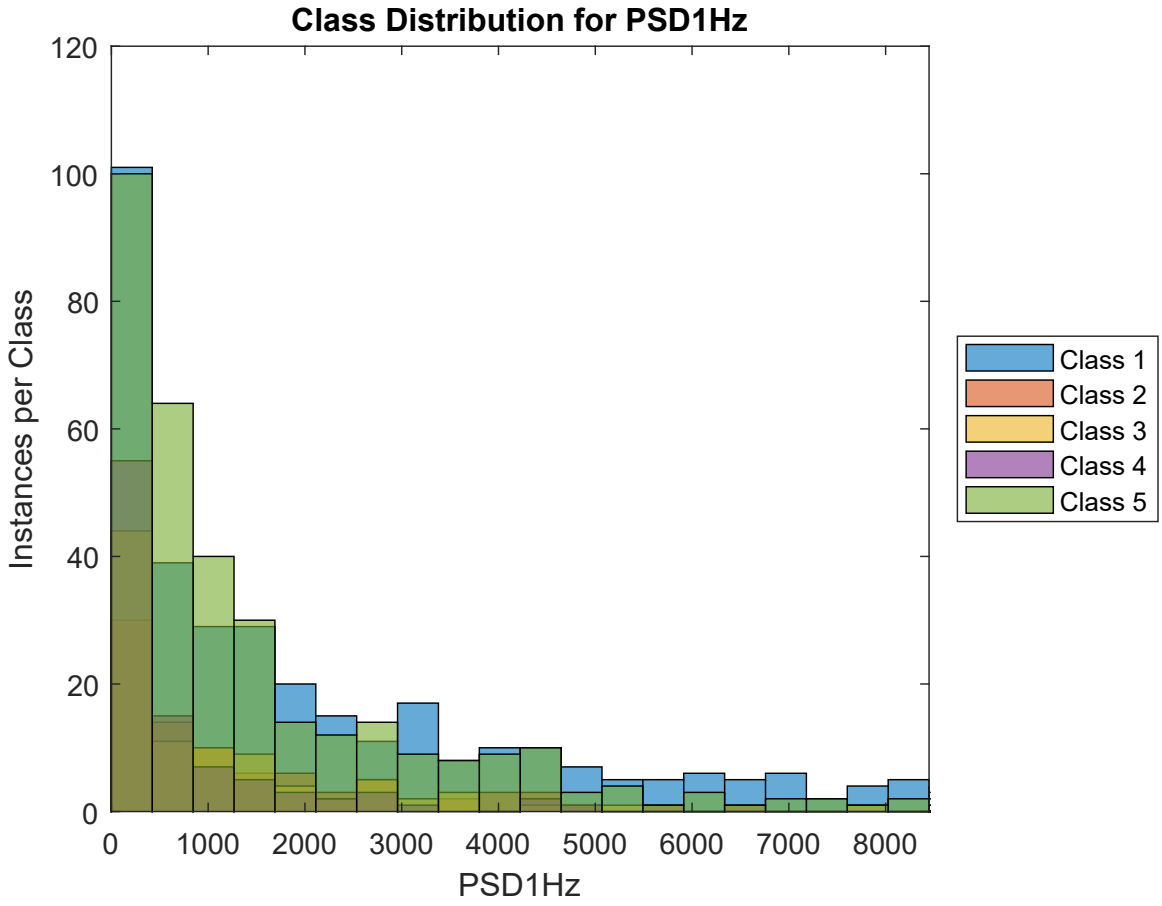

Figure 5.70: Histogram of the class distribution of the PSD1Hz feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.397	0.344	0.284	0.405
	Class 2	0.397	-	0.436	0.356	0.482
	Class 3	0.344	0.436	-	0.410	0.412
	Class 4	0.284	0.356	0.410	-	0.317
	Class 5	0.405	0.482	0.412	0.317	-

Table 5.62: Area under the curve for class versus class performance of the feature PSD1Hz.

The PSD1Hz feature is the power spectral density from the Welch periodogram at a frequency of 1Hz. A histogram depicting the class distribution of the feature is shown in Figure 5.70. According to the histogram, the feature does not provide any noticeable discriminability between classes. The table of AUC values in Table 5.62 shows that the feature has moderate separability between classes 1 and 4, with extremely poor separability between other class combinations.

5.10.2 Power Spectral Density at 3Hz (PSD3Hz)

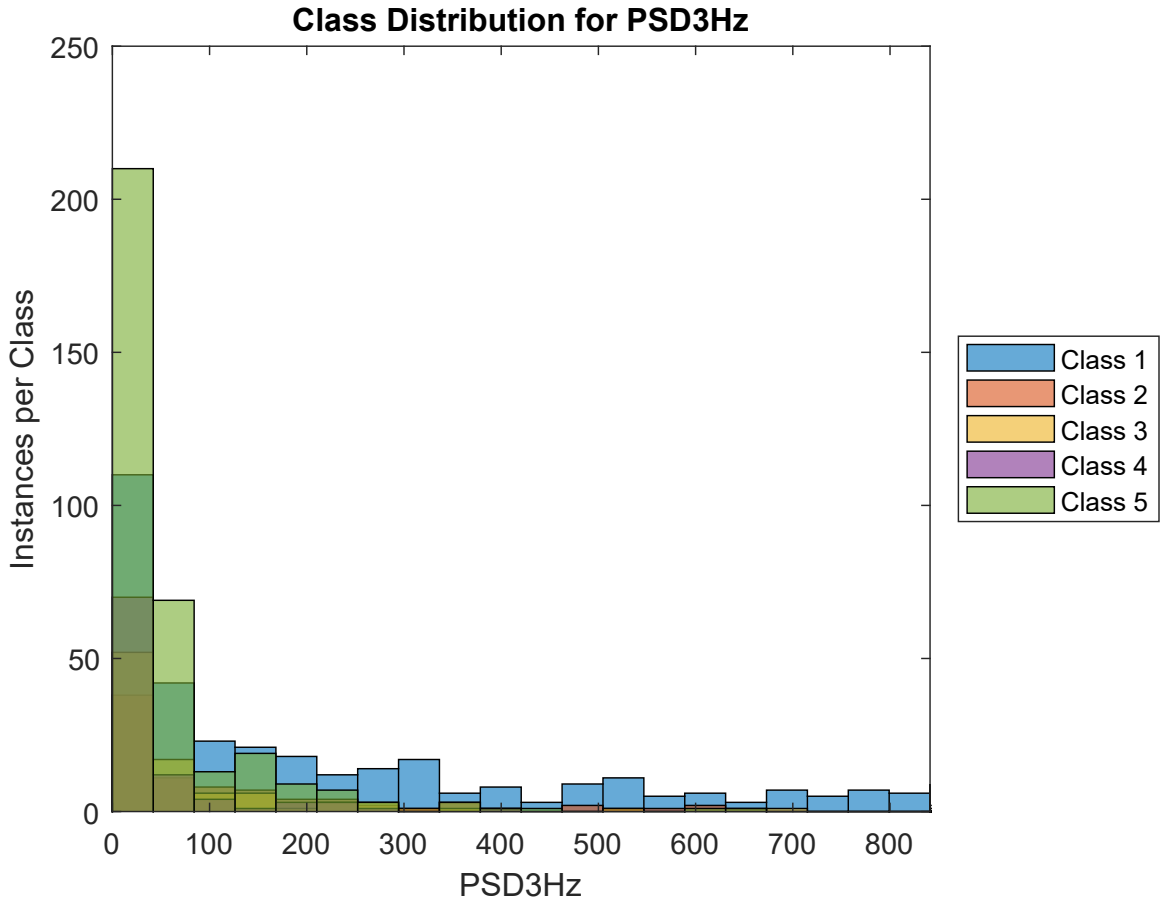

Figure 5.71: Histogram of the class distribution of the PSD3Hz feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.377	0.313	0.236	0.275
	Class 2	0.377	-	0.420	0.314	0.383
	Class 3	0.313	0.420	-	0.376	0.468
	Class 4	0.236	0.314	0.376	-	0.393
	Class 5	0.275	0.383	0.468	0.393	-

Table 5.63: Area under the curve for class versus class performance of the feature PSD3Hz.

The PSD3Hz feature is the power spectral density from the Welch periodogram at a frequency of 3Hz. A histogram depicting the class distribution of the feature is shown in Figure 5.71. According to the histogram, the feature does not provide any noticeable discriminability between classes 2, 3 and 4. The table of AUC values in Table 5.63 shows that the feature has moderate separability between classes 1 and 5, and classes 1 and 4, with extremely poor separability between other class combinations.

5.10.3 Power Spectral Density at 5Hz (PSD5Hz)

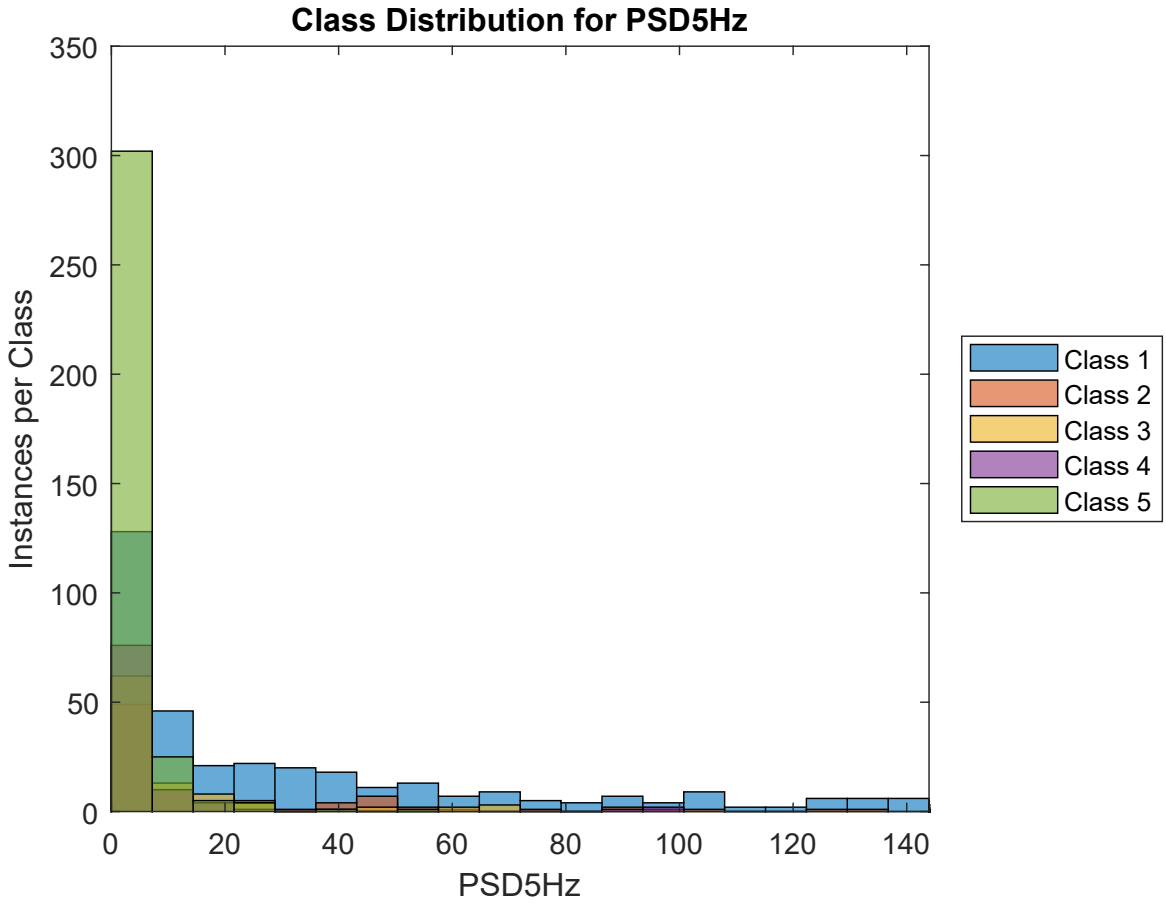

Figure 5.72: Histogram of the class distribution of the PSD5Hz feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.364	0.298	0.221	0.209
	Class 2	0.364	-	0.424	0.300	0.304
	Class 3	0.298	0.424	-	0.362	0.366
	Class 4	0.221	0.300	0.362	-	0.471
	Class 5	0.209	0.304	0.366	0.471	-

Table 5.64: Area under the curve for class versus class performance of the feature PSD5Hz.

The PSD5Hz feature is the power spectral density from the Welch periodogram at a frequency of 5Hz. A histogram depicting the class distribution of the feature is shown in Figure 5.72. It should be noted the majority of class 5 values are concentrated between 0 and 12W. According to the histogram, the feature does not provide any noticeable discriminability between classes 2, 3 and 4. The table of AUC values in Table 5.64 shows that the feature has moderate separability between classes 1 with classes 3, 4, and 5, with extremely poor separability between other class combinations.

5.10.4 Power Spectral Density at 7Hz (PSD7Hz)

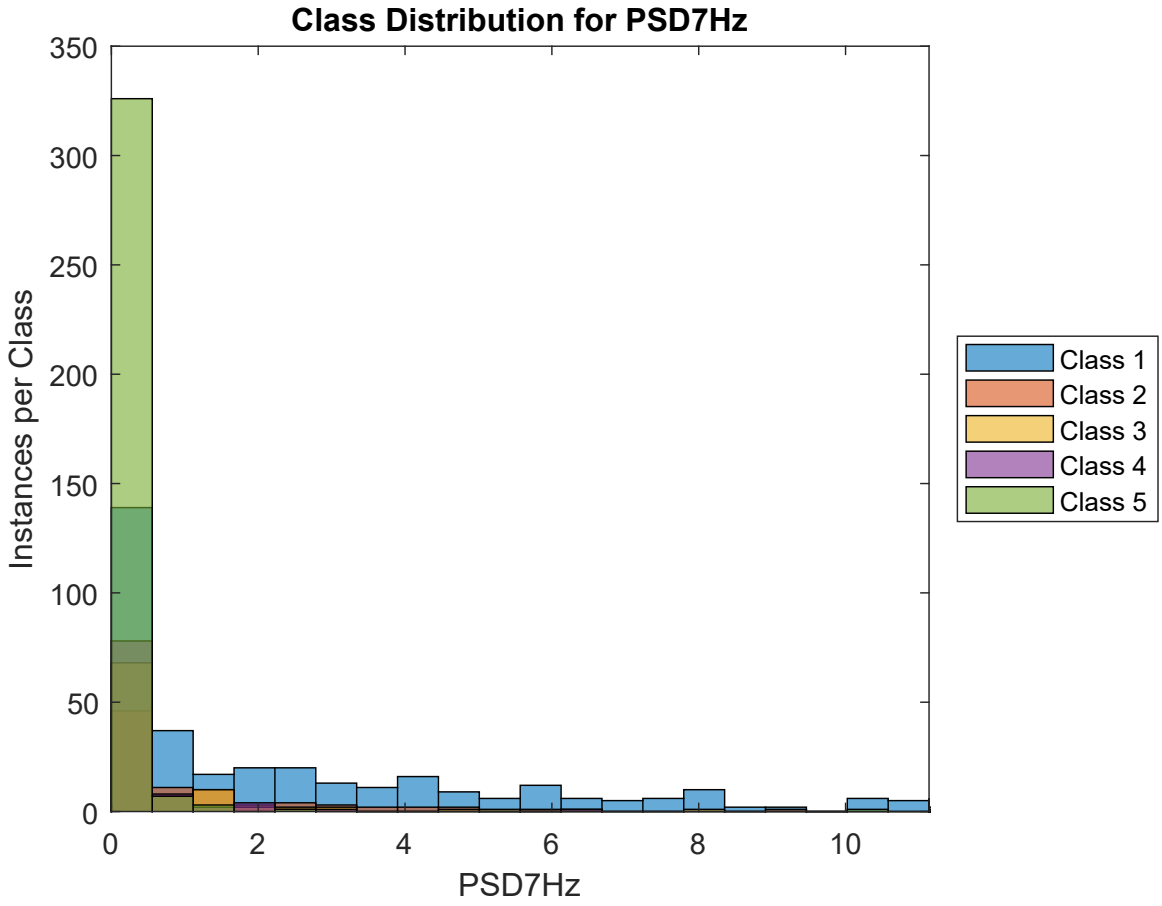

Figure 5.73: Histogram of the class distribution of the PSD7Hz feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.374	0.287	0.211	0.171
	Class 2	0.374	-	0.388	0.271	0.208
	Class 3	0.287	0.388	-	0.364	0.293
	Class 4	0.211	0.271	0.364	-	0.444
	Class 5	0.171	0.208	0.293	0.444	-

Table 5.65: Area under the curve for class versus class performance of the feature PSD7Hz.

The PSD7Hz feature is the power spectral density from the Welch periodogram at a frequency of 7Hz. A histogram depicting the class distribution of the feature is shown in Figure 5.73. It should be noted the majority of class 5 values are concentrated between 0 and 1W. The table of AUC values in Table 5.65 shows that the feature has good separability between classes 1 and 5, with moderate or poor separability for other class combinations.

5.10.5 Power Spectral Density at 9Hz (PSD9Hz)

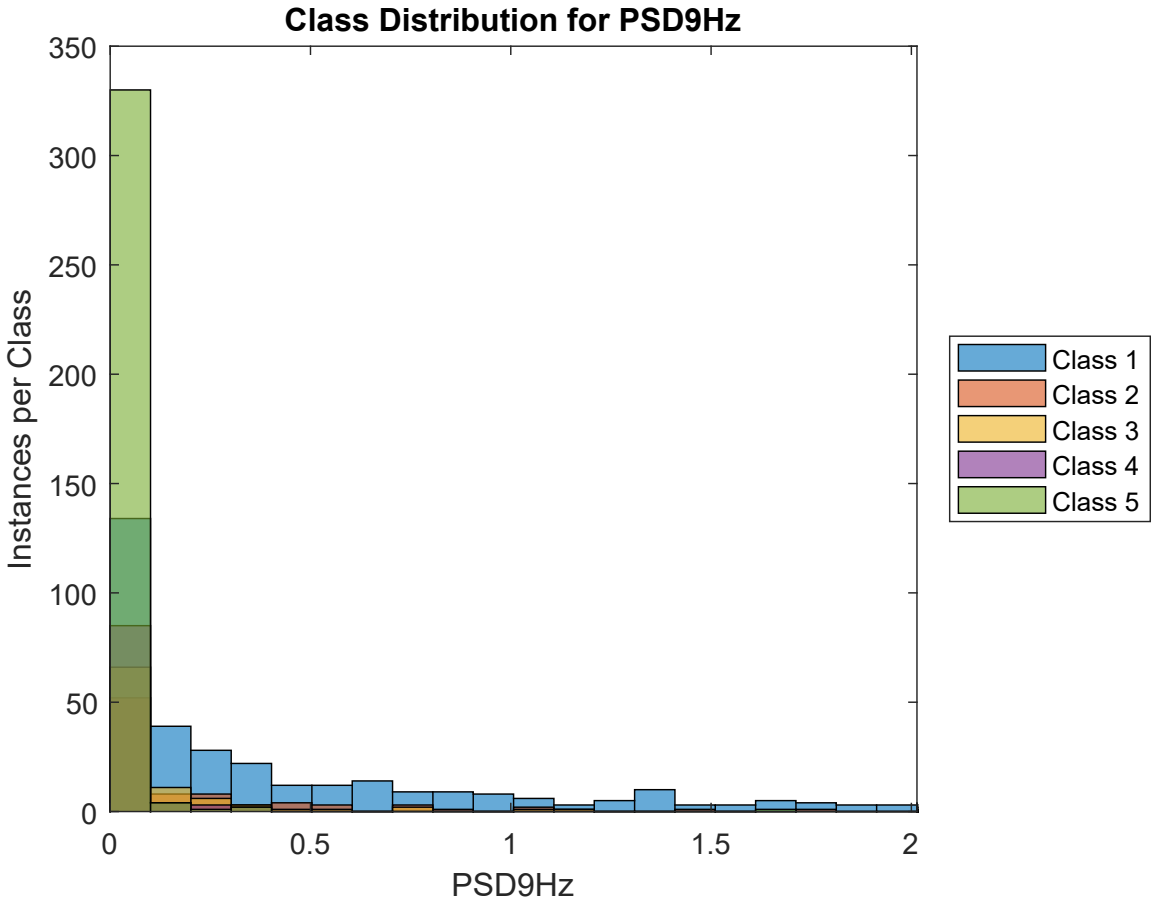

Figure 5.74: Histogram of the class distribution of the PSD9Hz feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.351	0.275	0.182	0.159
	Class 2	0.351	-	0.398	0.268	0.233
	Class 3	0.275	0.398	-	0.364	0.332
	Class 4	0.182	0.268	0.364	-	0.482
	Class 5	0.159	0.233	0.332	0.482	-

Table 5.66: Area under the curve for class versus class performance of the feature PSD9Hz.

The PSD9Hz feature is the power spectral density from the Welch periodogram at a frequency of 9Hz. A histogram depicting the class distribution of the feature is shown in Figure 5.74. It should be noted the majority of class 5 values are concentrated between 0 and 0.28W. The table of AUC values in Table 5.66 shows that the feature has good separability between classes 1 with classes 4 and 5, with moderate or poor separability for other class combinations. The feature provides no significant separability between classes 4 and 5.

5.10.6 Power Spectral Density at 13Hz (PSD13Hz)

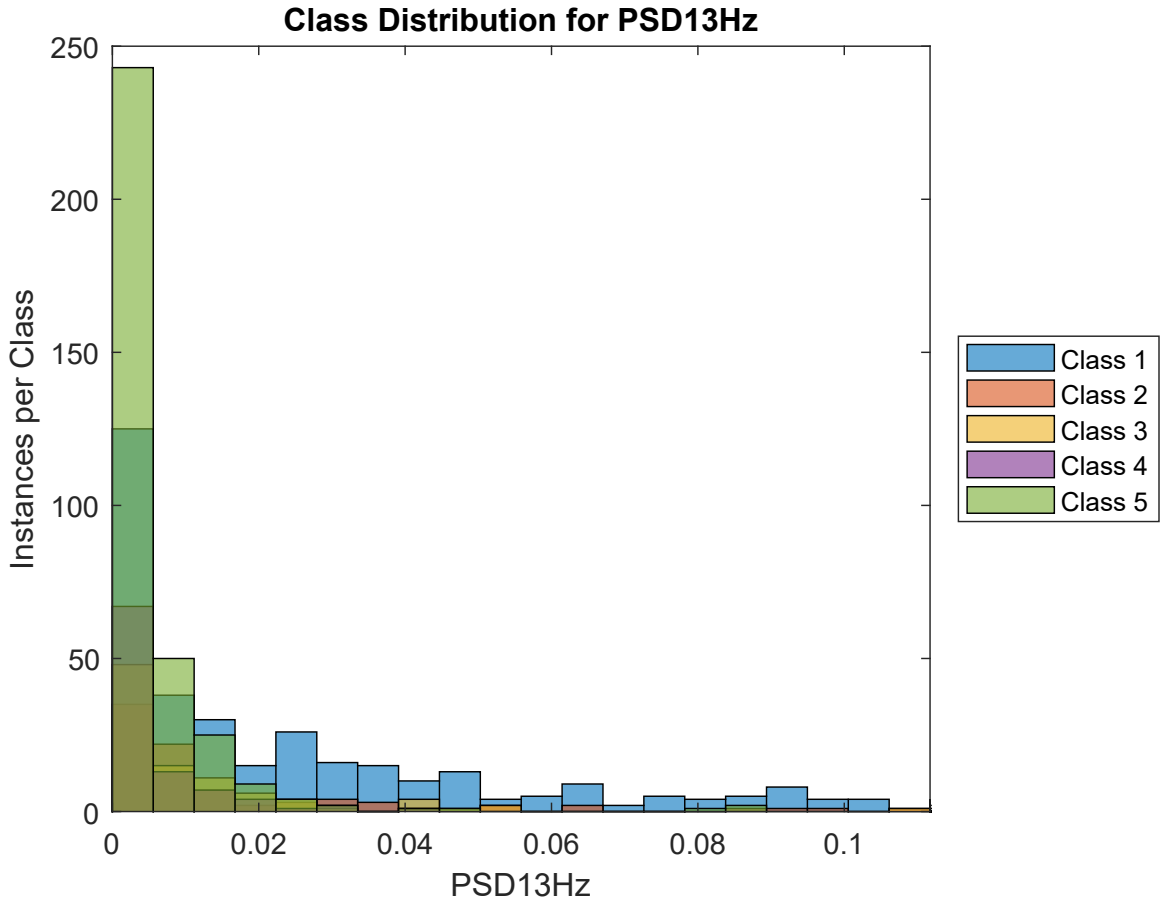

Figure 5.75: Histogram of the class distribution of the PSD13Hz feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.404	0.346	0.260	0.259
	Class 2	0.404	-	0.436	0.317	0.314
	Class 3	0.346	0.436	-	0.376	0.379
	Class 4	0.260	0.317	0.376	-	0.497
	Class 5	0.259	0.314	0.379	0.497	-

Table 5.67: Area under the curve for class versus class performance of the feature PSD13Hz.

The PSD13Hz feature is the power spectral density from the Welch periodogram at a frequency of 13Hz. A histogram depicting the class distribution of the feature is shown in Figure 5.75. According to the histogram, the feature does not provide any noticeable discriminability between classes 2, 3 and 4. The table of AUC values in Table 5.67 shows that the feature has moderate separability between classes 1 and 5, and classes 1 and 4, with extremely poor separability between other class combinations.

5.10.7 Power Spectral Density at 17Hz (PSD17Hz)

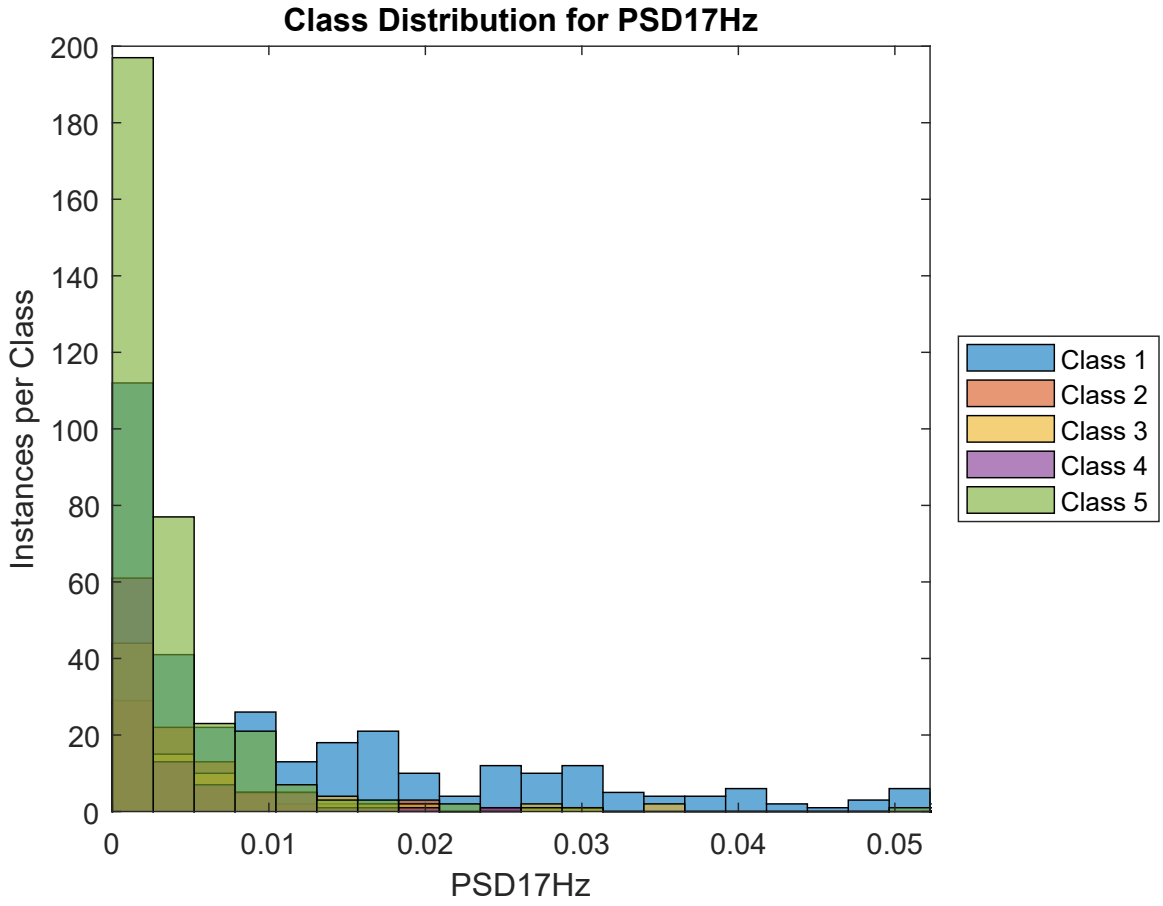

Figure 5.76: Histogram of the class distribution of the PSD17Hz feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.397	0.350	0.267	0.266
	Class 2	0.397	-	0.444	0.330	0.324
	Class 3	0.350	0.444	-	0.383	0.385
	Class 4	0.267	0.330	0.383	-	0.496
	Class 5	0.266	0.324	0.385	0.496	-

Table 5.68: Area under the curve for class versus class performance of the feature PSD17Hz.

The PSD17Hz feature is the power spectral density from the Welch periodogram at a frequency of 17Hz. A histogram depicting the class distribution of the feature is shown in Figure 5.76. According to the histogram, the feature does not provide any noticeable discriminability between classes 2, 3 and 4. The table of AUC values in Table 5.68 shows that the feature has moderate separability between classes 1 and 5, and classes 1 and 4, with extremely poor separability between other class combinations.

5.10.8 Power Spectral Density at 21Hz (PSD21Hz)

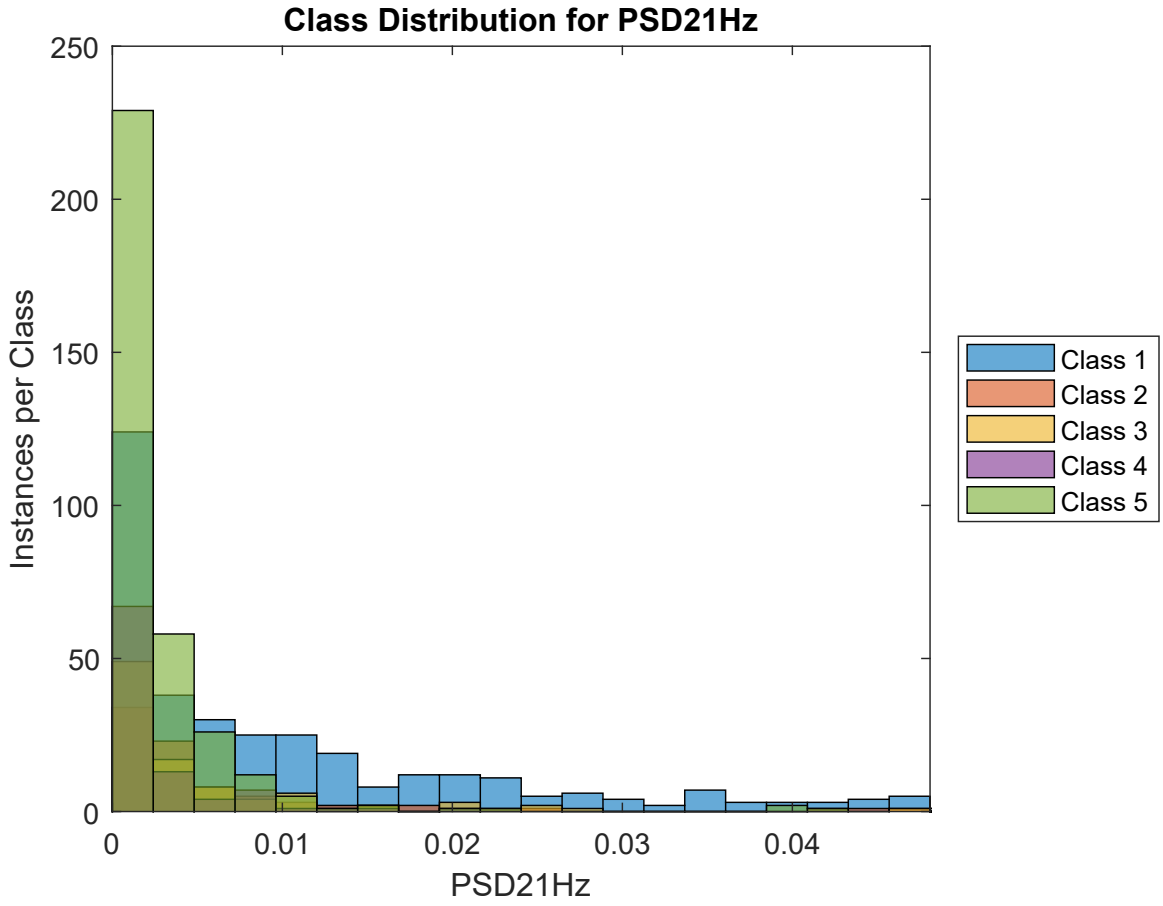

Figure 5.77: Histogram of the class distribution of the PSD21Hz feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.400	0.352	0.267	0.266
	Class 2	0.400	-	0.442	0.328	0.323
	Class 3	0.352	0.442	-	0.383	0.386
	Class 4	0.267	0.328	0.383	-	0.497
	Class 5	0.266	0.323	0.386	0.497	-

Table 5.69: Area under the curve for class versus class performance of the feature PSD21Hz.

The PSD21Hz feature is the power spectral density from the Welch periodogram at a frequency of 21Hz. A histogram depicting the class distribution of the feature is shown in Figure 5.77. According to the histogram, the feature does not provide any noticeable discriminability between classes 2, 3 and 4. The table of AUC values in Table 5.69 shows that the feature has moderate separability between classes 1 and 5, and classes 1 and 4, with extremely poor separability between other class combinations.

5.10.9 Power Spectral Density at 29Hz (PSD29Hz)

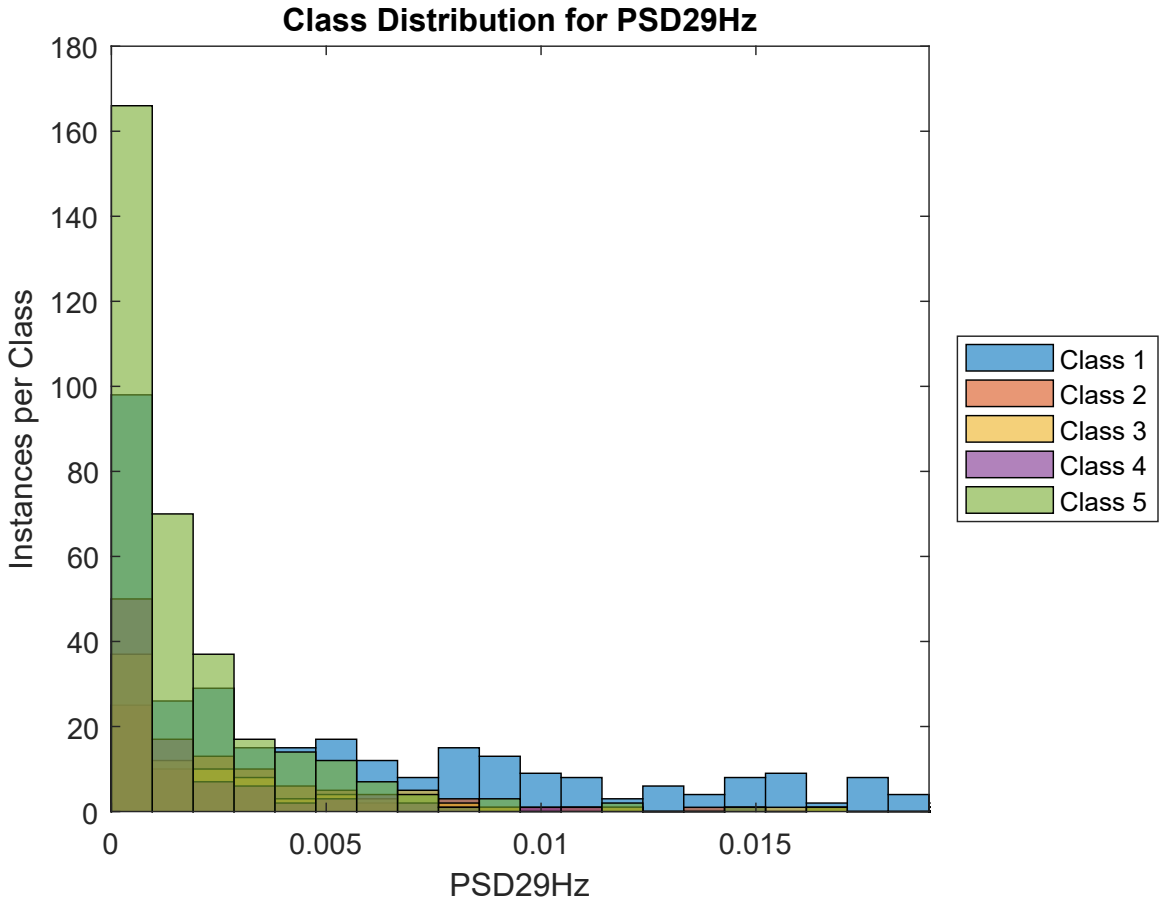

Figure 5.78: Histogram of the class distribution of the PSD29Hz feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.399	0.351	0.269	0.267
	Class 2	0.399	-	0.442	0.330	0.325
	Class 3	0.351	0.442	-	0.387	0.388
	Class 4	0.269	0.330	0.387	-	0.496
	Class 5	0.267	0.325	0.388	0.496	-

Table 5.70: Area under the curve for class versus class performance of the feature PSD29Hz.

The PSD29Hz feature is the power spectral density from the Welch periodogram at a frequency of 29Hz. A histogram depicting the class distribution of the feature is shown in Figure 5.78. According to the histogram, the feature does not provide any noticeable discriminability between classes 2, 3 and 4. The table of AUC values in Table 5.70 shows that the feature has moderate separability between classes 1 and 5, and classes 1 and 4, with extremely poor separability between other class combinations.

5.10.10 Relative Power

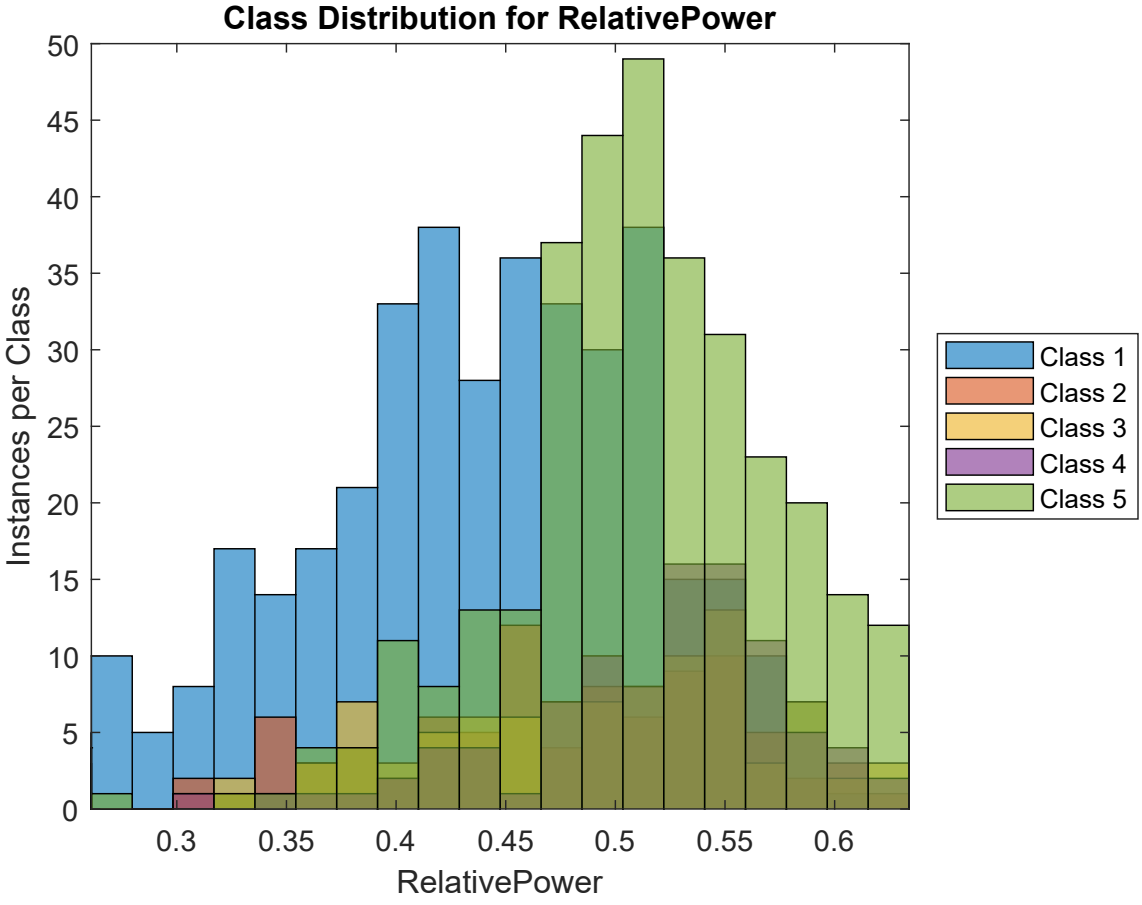

Figure 5.79: Histogram of the class distribution of the RelativePower feature.

		Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Signal Quality classes	**Class 1**	-	0.361	0.318	0.195	0.210
	Class 2	0.361	-	0.451	0.325	0.344
	Class 3	0.318	0.451	-	0.383	0.400
	Class 4	0.195	0.325	0.383	-	0.472
	Class 5	0.210	0.344	0.400	0.472	-

Table 5.71: Area under the curve for class versus class performance of the feature RelativePower.

The RelativePower feature was computed as the ratio between the total power of the Welch periodogram in the frequency range associated with good signal, 1-2.25Hz, and the frequency range from 0-8Hz, as shown in Equation 5.10. The periodogram for this feature was constructed differently than the previous PSD features. A window length of 1 second was used, with an overlap of 0.25 seconds. This feature was inspired from work done by Elgendi [28].

$$RelativePower = \frac{\sum_{f=1}^{2.25} PSD(f)}{\sum_{f=0}^{8} PSD(f)} \tag{5.10}$$

A histogram depicting the class distribution of the feature is shown in Figure 5.79. According to the histogram, there is moderate separability between classes 1 and 5. Additionally, the table of AUC values in Table 5.71 shows that the feature provides good separability between classes 1 and 4. Other class combinations provide poor separability.

Chapter 6

Feature Selection and Classification Results

6.1 Feature Selection

A total of 71 features were extracted for the analysis. A reduction in the number of features was desired to render the classification system as efficient as possible by eliminating poor and unnecessary features. However, the univariate analysis of the features indicated that no individual feature was adequate for discriminating between the five signal quality classes. Therefore, a multivariate approach was used to identify a subset of features, which together would provide the best class discrimination.

Using Weka [37], feature selection was done with the wrapper method, which performs various iterations with different subsets of features, and selects the subset offering the maximum class discrimination. The wrapper method is preferred to ranker methods as it considers feature dependencies and correlations for feature subset selection [38]. Feature dependency indicates whether multiple features together can provide good class discrimination where each feature individually would not. Un-

der an individual ranker approach, such dependencies would not be considered in the feature assessment [39]. Feature correlation is when multiple features are correlated with one another, indicating that they provide the same class discrimination information, thus making them redundant. Therefore, a wrapper method was chosen for the feature selection process.

Wrapper methods evaluate subsets using a classifier, providing a direct link between the feature selection and the classifiers used in the classification stage [38]. A random forest classifier was selected for use in this feature selection process. This was chosen as it was deemed to be sufficiently complex to provide good results for the features with poor univariate discriminability, while being simple enough not to be computationally intensive or time consuming.

To obtain the best results in a reasonable time, a greedy step-wise approach was used. While an exhaustive method is generally considered to be superior, it is computationally intensive and therefore time consuming for searches with a high number of features, as there are in this case. A drawback of the greedy step-wise approach is its susceptibility to getting stuck in a local optimum, rather than finding the overall optimum as the exhaustive search would.[39] To mitigate this, both forward and backwards greedy step-wise approaches were applied. Backwards approaches begin with all of the features, and attempt to remove individual features without compromising the discriminability of the feature subset, whereas forwards approaches begin with no features and work in reverse to identify the optimal subset. The performance of a particular subset is evaluate by Weka using a *merit score*, in a range from 0 (worst) to 1 (best), based on the classification accuracy. The backwards approach resulted in a subset of 70 features, with a merit score of 0.778. The forwards approach resulted in 9 features, with a merit score of 0.777. Due to the similarity in merit score for both approaches, the 9 feature subset from the forwards approach

was selected. Repetition of the forwards approach can be used to strengthen the confidence in the results, if repetitions produce the same subset of features. This forwards approach was repeated 3 times, with the same 9 features being selected at each repetition. The selected features are listed below:

- BillauerPeaks
- ZeroCrossings
- medianN
- medianR
- stdevE
- medianACC
- stdevACC
- ACPeakVals1
- ACPeakVals2

6.2 Classification Results

Five classifiers were evaluated using the annotated dataset and the nine features chosen from the feature selection process. Evaluations of each of the classifiers were done using a modified 13-fold cross validation training/testing scheme. This was done to maximize use of the annotated dataset, while ensuring that the classifier is not overfitting to the dataset. In 13-fold cross-validation, the dataset is divided into 13 subsets. At each iteration of the cross validation process, one subset is used as a testing set, while the remaining 12 subsets are used to train the classifier. This is repeated 13 times until classifier predictions have been made for each subset. Standard practice for cross-validation is to randomize the selection of subset membership, while ensuring class balance. However, due to concerns regarding overfitting, this process is modified. Instead, divisions are made to ensure all data segments from a particular participant are contained within the same subset. Thus, at any given iteration, the classifier is not training and testing on data from the same participant.

Confusions tables are constructed for each classifier comparing the actual classes with predicted classes of all instances. The overall accuracy is computed as the percentage of correctly classified instances, over the total number of instances.

Performance of the classifiers is also linked to the rater agreement from Section 4.4 using histograms to compare the correctness of the classifier with the rater agreement per instance (from Equation 4.1).

6.2.1 *k*-Nearest Neighbour

		Predicted Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Actual Signal Quality Classes	**Class 1**	233	37	30	24	72
	Class 2	42	7	9	13	19
	Class 3	29	12	6	13	35
	Class 4	19	9	11	13	43
	Class 5	64	19	32	47	176

Table 6.1: Confusion Matrix showing the the performance of the k-Nearest Neighbour classifier. Each element of the matrix shows the number of segments for each actual and predicted class combination. The overall accuracy of the classifier from the 1014 segments was 42.9%.

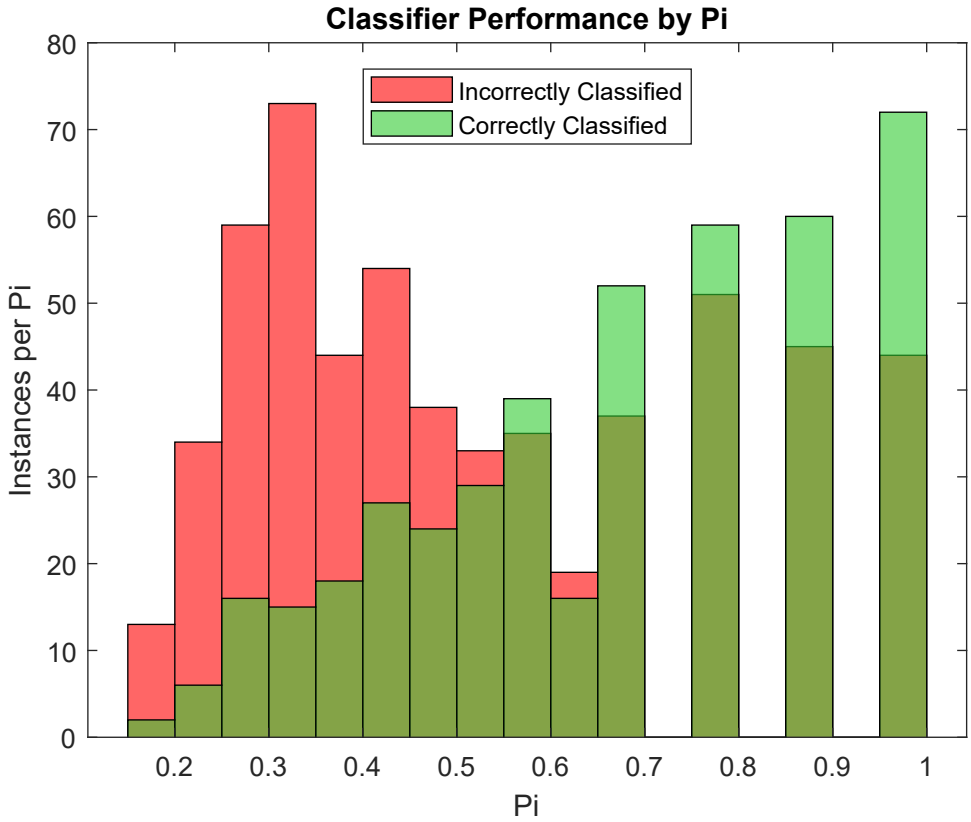

Figure 6.1: Performance of the k-Nearest Neighbour classifier compared to the rater agreement for each segment (instance).

The k-Nearest Neighbour (k-NN) classifier was evaluated as an initial, simple classifier. In this classifier, instances are classified based on the class of nearby instances in the feature space. Proximity of the instance to surrounding instances can be measured using various distance measures. [40] The classifier used in this case employed the Euclidean distance as its distance metric.

A simple, binary example of a k-NN is shown in Figure 6.2, where O and X represent two classes plotted in a two-feature plot. The instances, q1 and q2, are classified based on the class of the nearest neighbouring instances.

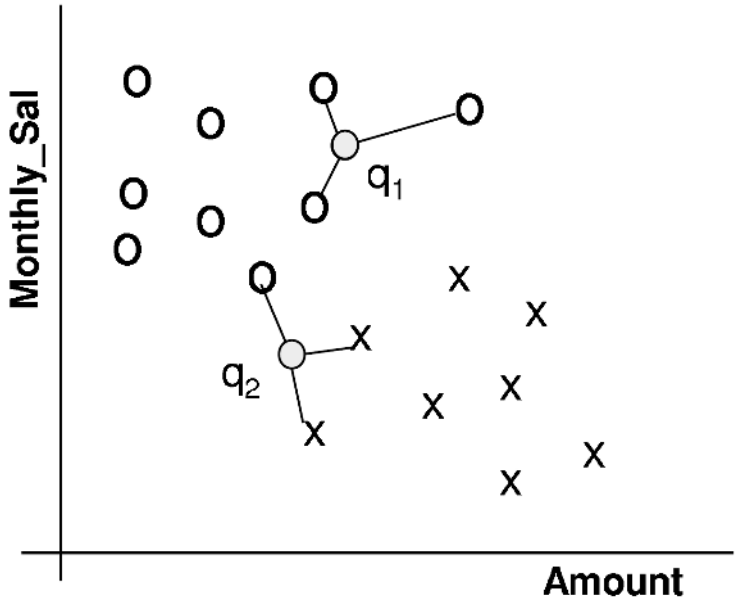

Figure 6.2: Example of simple, binary k-Nearest Neighbour classifier operation, where O and X denote different classes. Reproduced from [40].

Implementation of the classifier is done in Matlab using the *fitcknn* function. For our dataset, the k-NN classifier provided an accuracy of 42.9%. As seen in Table 6.1, the classifier performed poorly, even incorrectly classifying instances from extreme classes (class 1 misclassified as class 5, and vice-versa). Figure 6.1 shows that the classifier performed poorly, even for instances with high rater agreement. Thus, a more complex classifier is required for this problem.

6.2.2 Multi-Class Support Vector Machine

		Predicted Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Actual Signal Quality Classes	**Class 1**	370	22	1	2	1
	Class 2	62	16	3	6	3
	Class 3	27	42	8	14	4
	Class 4	4	22	19	45	5
	Class 5	1	17	30	288	2

Table 6.2: Confusion Matrix showing the the performance of the Multi-Class Support Vector Machine classifier. Each element of the matrix shows the number of segments for each actual and predicted class combination. The overall accuracy of the classifier from the 1014 segments was 43.5%.

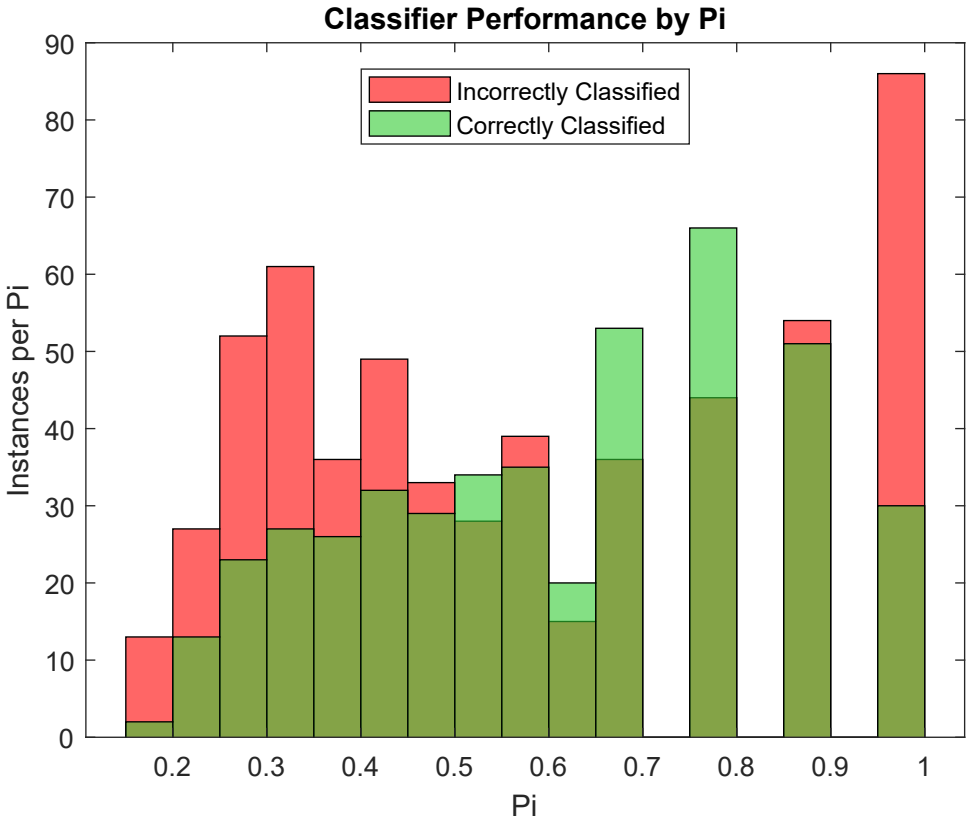

Figure 6.3: Performance of the Multi-Class Support Vector Machine classifier compared to the rater agreement for each segment (instance).

Support Vector Machines (SVMs) rely on support vectors to create a hyperplane in feature space to obtain maximum separation of classes. Each instance in the training set is represented as a multidimensional vector, called a support vector, with the dimensionality of the vector being the number of features. For data with high dimensionality, kernel functions are used to create complex hyperplanes, maximizing the distance between classes. [41]

SVMs are inherently binary classifiers. It was adapted to use in this multi-class problem by training a set of binary SVMs for each class versus the rest of the data. Thus, each SVM specializes in separating one class from the rest of the data. Each instance in the test set is then passed through the SVMs in order, one at a time (from the SVM for class 1 to SVM for class 5) until it is classified as belonging to the class associated with that SVM. This was implemented in Matlab using the *multisvm* function by Cody Neuburger, obtained from the Matlab Central File Exchange. [42]

The classifier provided an accuracy of 43.5%. As seen in Table 6.2, the classifier performed well for class 1, while performing poorly for all other classes. Figure 6.3 shows that the classifier performed poorly, even for instances with high rater agreement.

6.2.3 Naïve Bayes Classifier

		Predicted Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Actual Signal Quality Classes	**Class 1**	261	118	10	7	0
	Class 2	19	44	10	12	5
	Class 3	12	32	17	16	18
	Class 4	1	10	10	16	58
	Class 5	0	4	11	16	307

Table 6.3: Confusion Matrix showing the the performance of the Naïve Bayes classifier. Each element of the matrix shows the number of segments for each actual and predicted class combination. The overall accuracy of the classifier from the 1014 segments was 63.6%.

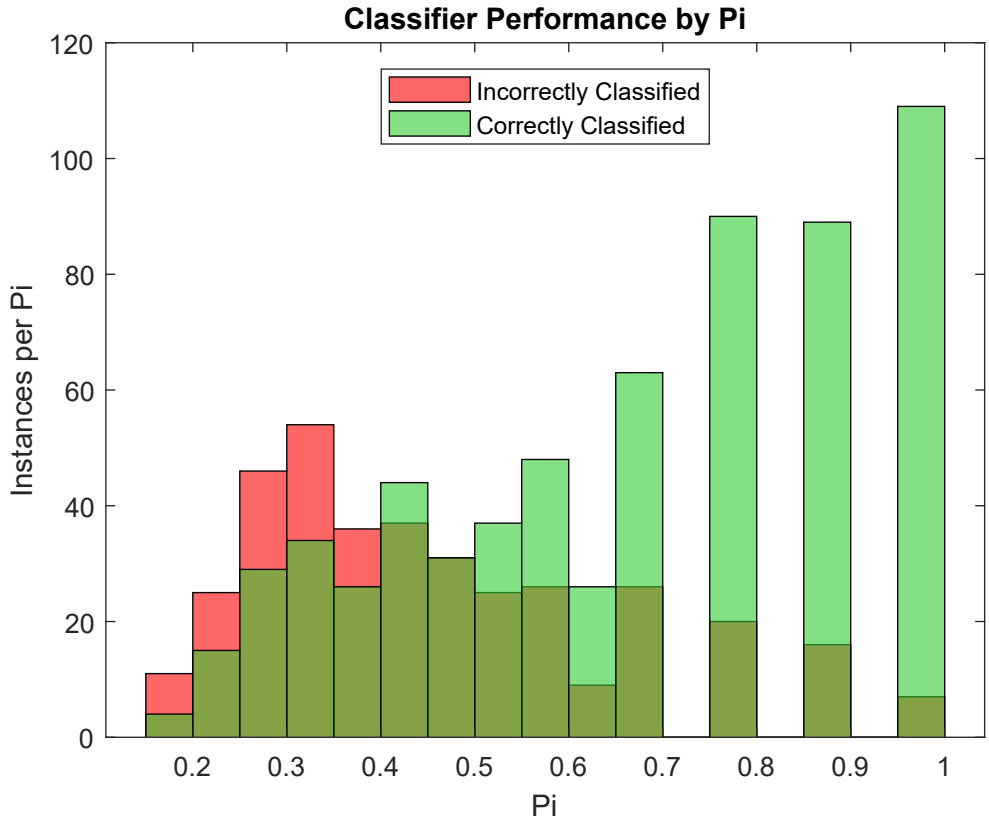

Figure 6.4: Performance of the Naïve Bayes classifier compared to the rater agreement for each segment (instance).

The Naïve Bayes classifier uses Bayes'rule to determine the posterior probability of an instance belonging to a particular class based on the class probability distributions in the feature space. This classifier operates under the assumption that the features are independent of one another, given the class. At each instance, given the feature vector for the instance, the posterior probability is computed for each class. The class yielding the maximum posterior probability is chosen as the predicted class of the instance. [43]

Implementation of the classifier is done in Matlab using the *fitNaiveBayes* function. Classification results for the Naïve Bayes classifier are shown in Table 6.3. Overall accuracy of the classifier is 63.6%. The classifier's performance for classes 3 and 4 is especially poor, with many more instances belonging to those classes classified incorrectly than correctly. Over one third of class 1 instances were predicted as class 2. The best classifier performance was obtained with class 5. However, it ought to be noted that classifier performance was significantly better for instances with higher rater agreement, as depicted in Figure 6.4.

6.2.4 Decision Tree

		Predicted Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Actual Signal Quality Classes	**Class 1**	306	62	17	10	1
	Class 2	48	19	14	7	2
	Class 3	22	25	28	15	5
	Class 4	6	7	22	32	28
	Class 5	0	2	14	29	293

Table 6.4: Confusion Matrix showing the the performance of the Decision Tree classifier. Each element of the matrix shows the number of segments for each actual and predicted class combination. The overall accuracy of the classifier from the 1014 segments was 66.9%.

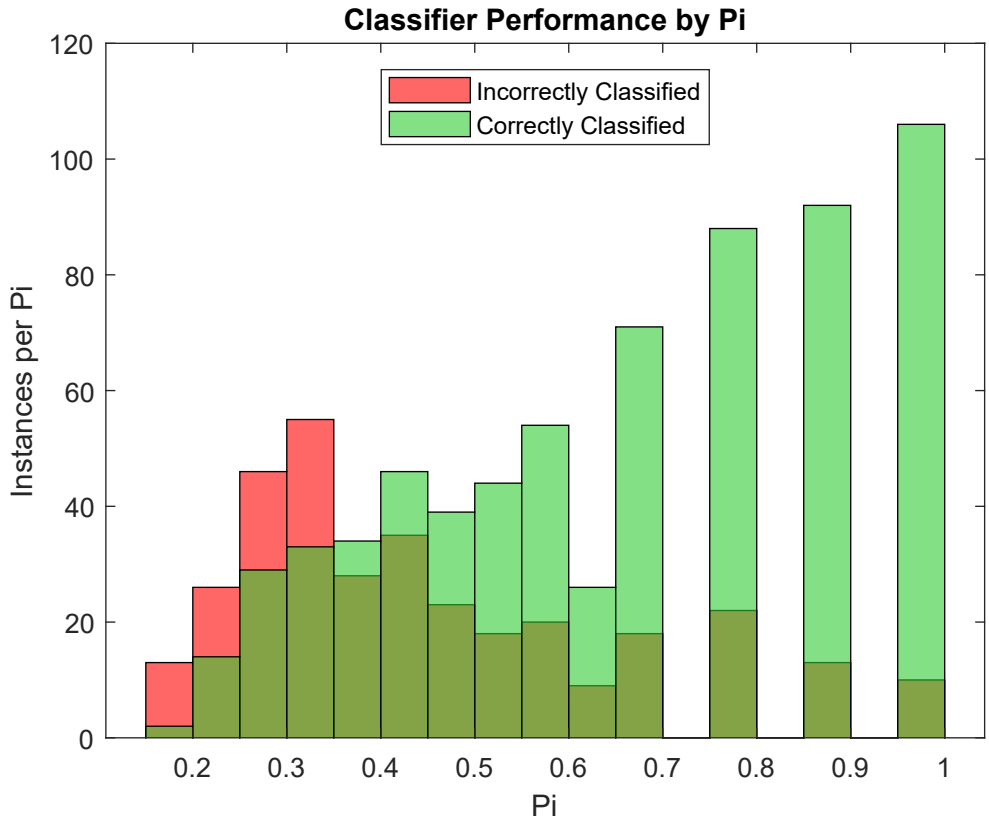

Figure 6.5: Performance of the Decision Tree classifier compared to the rater agreement for each segment (instance).

A Decision Tree is a hierarchical feature-based decision making structure consisting of three types of nodes; root nodes, internal nodes, and leaf nodes. The leaf nodes are assigned to specific classes, in this case the five quality classes. The root nodes and internal nodes are used for feature-based rules and test conditions to separate the instances based on characteristics relative to the features. [44] [45]

A simple example of the structure of a Decision Tree is shown in Figure 6.6. In the example, class decisions regarding whether animals are mammals are made using test conditions based on features such as 'body temperature' and 'gives birth'.

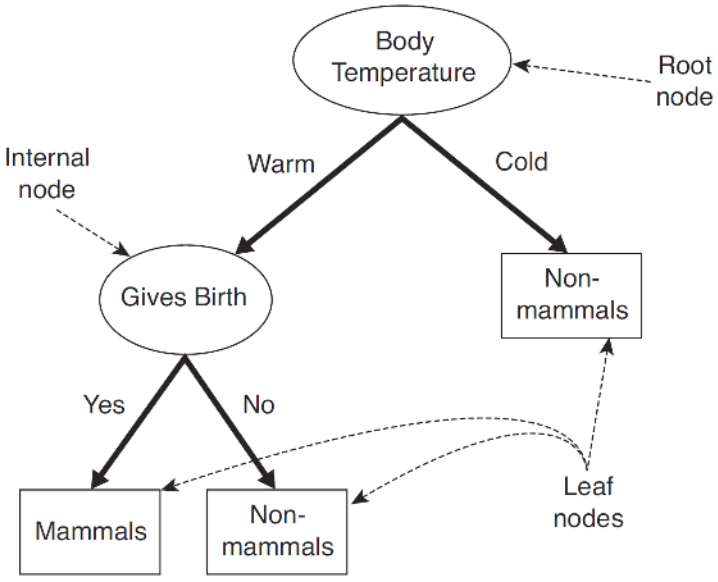

Figure 6.6: Example of simple, binary Decision Tree classifier structure. Reproduced from [44].

Implementation of the classifier is done in Matlab using the *fitctree* function. Classification results for the Decision Tree classifier are shown in Table 6.4. Overall accuracy of the classifier is 66.9%. Classification for classes 2 and 3 are especially poor, though most of the predictions for class 2 are within one class from the actual (in the class 1 to class 3 range). However, the classifier performance was significantly better for instances with higher rater agreement, as depicted in Figure 6.5.

6.2.5 Random Forest

		Predicted Signal Quality Classes				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Actual Signal Quality Classes	**Class 1**	365	13	11	4	3
	Class 2	60	6	16	8	0
	Class 3	31	19	23	17	5
	Class 4	9	4	13	38	31
	Class 5	1	0	3	11	323

Table 6.5: Confusion Matrix showing the the performance of the Random Forest classifier. Each element of the matrix shows the number of segments for each actual and predicted class combination. The overall accuracy of the classifier from the 1014 segments was 74.5%.

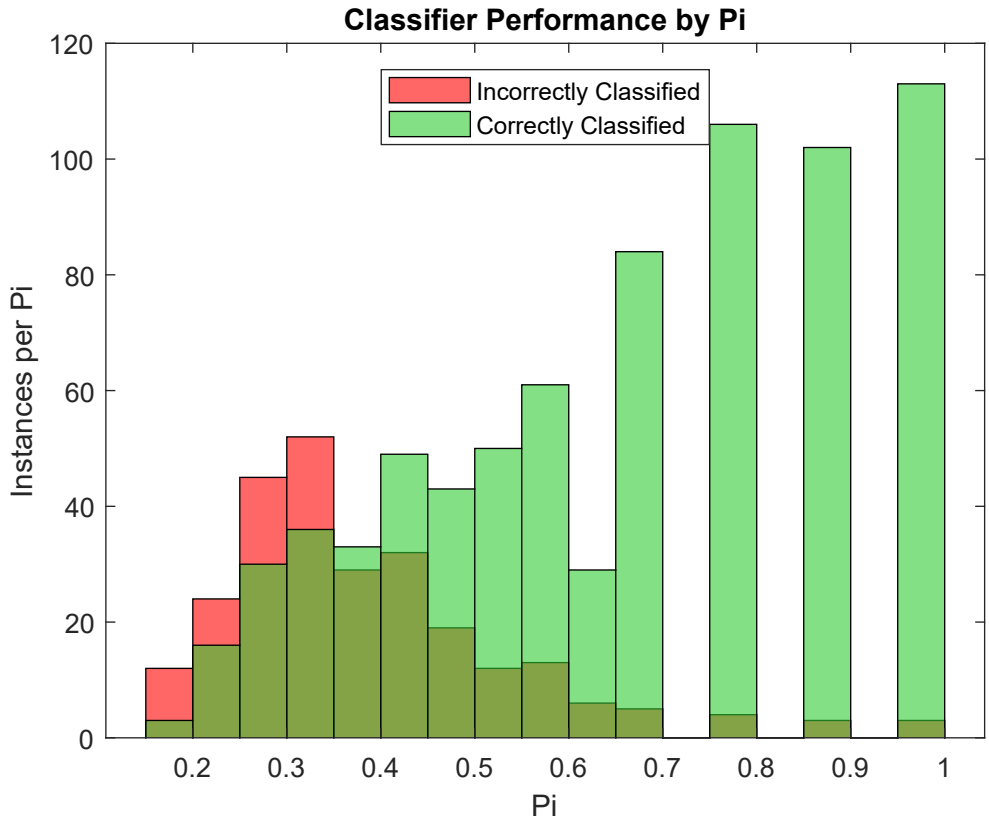

Figure 6.7: Performance of the Random Forest classifier compared to the rater agreement for each segment (instance).

An ensemble of Decision Tree classifiers can be used together in a voting meta-learning system to improve the classification accuracy. Individual trees in the ensemble are trained on a randomly selected subset of features, and each subset of features is chosen independently of the other subsets. This is known as a random forest classifier. [46] Implementation of the classifier in Matlab was done using the *Treebagger* function.

Classification results for the Decision Tree classifier are shown in Table 6.5. Overall accuracy of the classifier is 74.5%. Classification for classes 2 and 3 are especially poor, though most of the predictions for class 2 are within one class from the actual (in the class 1 to class 3 range). However, it ought to be noted that most classification errors occurred for instances with lower rater agreement of the true class, as depicted in Figure 6.7.

6.2.6 Classifier Selection

Evaluation of the classifiers yielded various accuracies, depicted in Table 6.6. The highest accuracy, 74.5%, was obtained from the Random Forest classifier, hence it was selected for the final algorithm. While this is the best accuracy obtained from the classifier analysis, it performs poorly for certain classes. However, the classes for which the classifier performed poorly appear to have low prevalence, thereby allowing an accuracy of 74.5%.

Random Forest is a meta-learning based approach using multiple decision tree classifiers. Each tree is trained using a randomly selected subset of instances and features. This allows each tree to specialise in discriminating between different subsets of data. The final classification for an instance (a 10-second segment) is obtained through a plurality voting system, in which each of the decision trees cast a vote. Thus, the weaknesses of one decision tree can be compensated by the others in the forest. This meta-learning approach is likely responsible for the higher accuracy obtained by this classifier.

Classifier	**Accuracy**
k-Nearest Neighbour	42.9%
Multi-Class SVM	43.5%
Naïve Bayes	63.6%
Decision Tree	66.9%
Random Forest	**74.5%**

Table 6.6: Accuracy of classifiers

Ultimately, the performance of the classifier is a reflection of the class discriminability of the features. The feature extraction process can be expanded to improve the performance of the algorithm, by searching for features with greater class discrimination, especially for the classes for which the classifier performed poorly. Classification methods with greater complexity can also be analyzed, however, it should

be noted that greater computational complexity of the algorithm will complicate its implementation on wrist-based devices.

Adjustments to the classification system may be required for use in atrial fibrillation (AF) monitoring, as the presence of AF itself may be registered as noise. Conclusive assessment of the effect of AF on the performance of the algorithm cannot be done without evaluating the algorithm with ambulatory data acquired from individuals suffering from AF. However, it is believed that features relying on the periodicity of noise-free data for class discrimination, such as those obtained from the correlogram, would be adversely affected. Meanwhile, features derived from the accelerometer signals are expected to be least affected, as they rely on the movement of the wrist, which is unlikely to be linked to an AF episode, to discriminate between classes.

Chapter 7

Overall Signal Quality

The signal quality classification algorithm developed in Chapter 6 is applied to the full dataset to evaluate the quality of wrist-based PPG throughout a 24-hour period, during ambulatory use.

7.1 Quality Assessment Procedure

The *Random Forest* classifier was selected as the best choice for the signal quality based on the work done in the previous section. The classifier is then used to provide signal quality analysis for the full 24 hours of data collected from each participant.

In Chapter 6, a modified 13-fold cross-validation scheme was used for the purposes of evaluating the classifier performance without overfitting. To ensure generalizability, the subset divisions were made to ensure that the classifier was not training and testing on data from the same participant within any given iteration. However, for this 24-hour assessment, accuracy of the results for this dataset are prioritized above generalizability of the classifier. To that end, this classifier is trained on the entire annotated dataset from Chapter 4, allowing it to be trained on data from each

participant, thereby maximizing its effectiveness for the full dataset.

Although the goal of the initial data collection (described in Chapter 3) was to acquire 24 hours of data from each participant, due to the practical limitations of scheduling, participant data collection was only approximately 24 hours. For participants whose data collection was less than 24 hours, their entire data was used. For participants for whom more than 24 hours of data was collected, only the first 24 hours of data was used. In some cases participants were required to remove the PPG wristbands themselves after 24 hours, and were unable to power down the device, thus, the device kept collecting data. Using only the first 24 hours of data ensures that data collected while the device was not worn is not included. This selected data was divided into continuous, 10 second, non-overlapping segments. The 9 features identified from the feature selection process were computed for all data segments. Based on these features, the Random Forest classifier provided signal quality ratings for each data segment.

7.2 Quality Assessment Results

	Signal Quality				
	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Non-Elderly	44.4%	6.2%	7.2%	9.8%	32.5%
Elderly	43.9%	3.9%	6.4%	7.3%	38.6%
Overall	44.2%	5.3%	6.9%	8.9%	34.8%

Table 7.1: Percentage of data belonging to each signal quality class. The signal quality ratings are provided by the trained random forest classifier.

The overall quality ratings for participants are compiled and displayed in Table 7.1. Overall, only 34.8% of the data belonged to class 5, which represents the noise-free quality class. While this is a small proportion of the total data, 34.8% the quality

of the signal for that portion is clean enough to be used for cardiovascular analysis. For both non-elderly and elderly participants, a plurality of the data belonged to class 1, representing the most noise corrupted data. This supports the assumption that low signal quality is caused by movement, as the device is located on the wrist, a site which experiences significant motion throughout daily activities. Class 5 data is more prevalent among the elderly demographic, with 38.6% of the data belonging to class 5, compared to only 32.5% for the non-elderly demographic. While mobility was an inclusion criteria for the selection of participants, differences in activity level between the two demographics are likely responsible for this disparity in data quality. However, this disparity of 6.1 percentage points is less than expected between the demographics. Also, the proportion of class 1 data between the groups is within one percentage point. It was expected that elderly individuals would engage in fewer, less motion intensive activities throughout the day. Thus, a class 5 percentage of over 50% was expected for the elderly demographic. Based on the results, however, it may be postulated that even hand movements from simpler daily activities are sufficient to cause corruption of the PPG signal. Overall, this indicates that wrist-based PPG may not be suitable for continuous monitoring is an ambulatory daily use setting, which is desired for the detection of episodic cardiovascular illnesses such as paroxysmal atrial fibrillation.

However, another 15.8% of the PPG data, which belonged to either classes 3 or 4, and could potentially be cleaned sufficiently to derive cardiovascular parameters, thereby enabling 50.6% of the data collected in an ambulatory setting to be useful.

7.2.1 Quality by Time of Day

To enable a more extensive analysis of the results, the data is subdivided by time of day for both non-elderly and elderly demographic groups, displayed in Tables 7.2

		Signal Quality Rating for Non-Elderly				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Time of Day	**Night (22h-7h)**	19.8%	3.6%	5.6%	10.0%	61.1%
	Morning (7h-12h)	48.8%	6.9%	8.2%	11.1%	25.0%
	Afternoon (12h-18h)	64.3%	8.1%	8.4%	9.1%	10.2%
	Evening (18h-22h)	64.6%	8.1%	7.8%	8.8%	10.7%

Table 7.2: Percentage of data from the non-elderly demographic belonging to each signal quality class for each time of day segment. The signal quality ratings are provided by the trained random forest classifier.

		Signal Quality Rating for Elderly				
		Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Time of Day	**Night (22h-7h)**	25.1%	2.9%	5.4%	7.4%	59.2%
	Morning (7h-12h)	56.4%	4.1%	6.5%	7.3%	25.8%
	Afternoon (12h-18h)	56.3%	4.3%	6.8%	7.5%	25.1%
	Evening (18h-22h)	52.1%	5.2%	7.8%	6.9%	28.0%

Table 7.3: Percentage of data from the elderly demographic belonging to each signal quality class for each time of day segment. The signal quality ratings are provided by the trained random forest classifier.

and 7.3, respectively. The choice of these specific subdivisions are based previous work [1]. A visualization of the signal quality over 24 hours is shown in Figure 7.1. This is an area plot in which the percentage of data belonging to each class, averaged for each 10-minute, non-overlapping time interval, is depicted in a vertically stacked manner. The area occupied by each colour represents the percentage occurrence of its associated class. The plot starts at beginning of the night period, at 22h (10:00pm), and continues for a full 24 hours.

Best data quality was obtained for both demographics during the night period, ranging from 22h (10:00pm) to 7h (7:00am), which is associated with lower levels of activity as the participants are likely asleep. Also, lower levels of activity are expected at times immediately preceding and succeeding sleep. The markers in Figure 7.1 identify the only time period for which class 5 data consists of more than 50% of the data. This period is seven hours for the non-elderly demographic, and eight hours

and ten minutes for the elderly demographic, though it ought to be noted that the elderly demographic has a brief spike in class 1 data at around 7h. Regardless, this indicates that the elderly demographic has a longer period during which high quality data is available. However, Tables 7.2 and 7.3 show a higher rate of class 5 data for non-elderly individuals at night. Thus, even though non-elderly individuals have a shorter period during which high quality data is available, a greater percentage for the data during that period belongs to class 5, relative to the elderly demographic. Likely, this is an indication that elderly individuals on average have a longer sleep duration, though may wake up at night more often, thereby resulting is lower rates of class 5 data. Unfortunately, the individual sleep cycles of the participants are unknown, therefore this hypothesis cannot be confirmed. Future work ought to include an activity log in which participants record their sleep times.

A notable difference between the two demographics in the portion of class 5 data during the afternoon and evening periods is seen in the visualization in Figure 7.1. The elderly demographic has a class 5 portion 14.9 and 17.3 percentage points higher than the non-elderly category, for the afternoon and evening time periods, respectively. This is likely due to the relatively lower activity levels expected among the elderly demographic. Despite this, only slightly over a quarter of their data during these periods is class 5. Thus, even for elderly individuals, wrist-based PPG technology may not be suitable for continuous monitoring during the day.

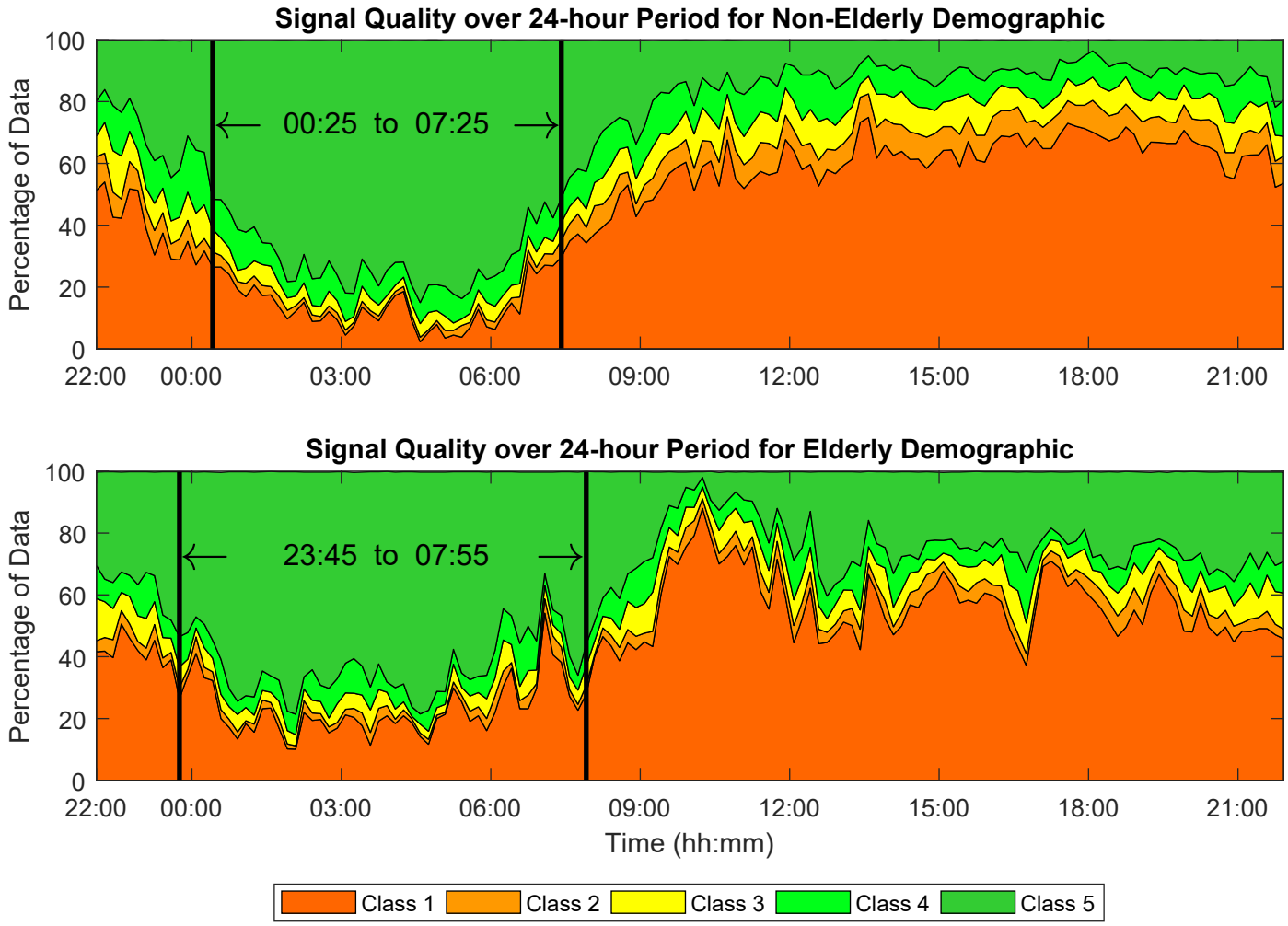

Figure 7.1: Quality of the PPG over the course of 24 hours for non-elderly and elderly demographics. The data in the plot is averaged to every ten minutes. The vertical lines delineate the only zones in which class 5 data consists of over 50% of the data.

Chapter 8

Conclusion

Signal quality of wrist-based photoplethysmography (PPG) technology was evaluated, as a potential tool for continuous, non-invasive cardiovascular monitoring. While wrist-based PPG devices such as the Empatica E4 are considered to be more user-friendly than electrocardiogram Holter monitors, they are also highly susceptible signal quality reduction due to motion artifacts. A dataset of 24-hour PPG was collected from 26 participants in an ambulatory setting. A subset of this dataset was annotated for signal quality by 17 raters, then used for developing a quality classification algorithm. This algorithm was then applied to the full dataset to evaluate wrist-based PPG over a 24-hour period.

Data itself was collected from 26 participants, 16 of whom were non-elderly (under 65 years of age), and 10 of whom were elderly (over 65 years of age). This was done over a period of approximately 24 hours for each participant. The PPG and accelerometer signals used in the analysis were acquired from the Empatica E4 wristband.

Reliability of the gold standard used for the training and evaluating the classifier was assessed. The only gold standard for signal quality of wrist-based PPG is visual

assessment of the signal, which can be subjective. To address this concern, a subset of the signal was annotated by 17 raters, with the mode used to produce the gold standard. The agreement between the raters obtained using Fleiss' Kappa was $\kappa = 0.4605$, which is interpreted as moderate agreement according to the benchmarks established by Landis and Koch [27]. The segments of signal used to establish the gold standard were extracted randomly from the dataset of each participant.

Classification of the signal quality was done using a machine learning based, *Random Forest* classifier. An analysis of various metrics derived from the PPG and accelerometer signals from the wrist concluded that a univariate system would be insufficient to discriminate between the classes. Consequently, a multivariate was necessary. Selection of the optimal subset of features (metrics derived from the signals) for class discrimination was done using greedy step-wise wrapper approach. Various classifiers were investigated, with the *Random Forest* classifier providing the best accuracy for discriminating between the quality classes. Overall, this classifier provided an accuracy of 74.5%.

Evaluation of wrist-based PPG technology for continuous, ambulatory use established that the technology provides high quality (class 5) data for only 34.8% of the day, on average. This in itself may not be sufficient for the detection of cardiovascular illnesses, however, time analysis of the signal quality reveals that the signal quality improves considerably during periods associated with sleep. Thus, while wrist-based PPG may be unsuitable for continuous monitoring during daily use, it has to the potential to be used in more limited applications.

Owing to its user-friendliness compared to ECG Holter monitors, wrist-based PPG has potential for long-term monitoring. However, the results form this study show that signal quality is insufficient in cases where continuous monitoring is desired. Thus, the device may be not be a suitable replacement to the ECG Holter monitor,

but rather a supplement for long-term cardiovascular monitoring.

8.1 Limitations

There are various limitations with the methodology of the study with regards to the subjects and the data.

The study recruited only mobile subjects, as the scope of the research was to ascertain whether the technology could be used during daily activities. Thus, a segment of the elderly population which is not mobile, was not included in the study. As the signal quality is largely associated with the amount of movement, the technology would be expected to provide better results in a non-mobile population. However, this was not assessed in this thesis and ought to be included in future work.

The data collection procedure itself involved did not last exactly 24 hours for each participant. This was largely due to the challenges of scheduling participant availabilities, in which researchers were unable to meet with some off-campus participants exactly 24 hours after the start of data collection. In some cases, the participants had to remove the devices themselves due to alterations in their scheduling, in which case they were sometimes unable to power off the devices, hence the devices recorded for hours after the designated 24 hours. To deal with this, only the first 24 hours of data was considered.

The data segments selected for the establishment of the gold standard were chosen randomly from each participant, with 39 ten second segments being chosen per participant. This random selection resulted in a class imbalance in the data set which was ultimately used in the training and testing of the classifier. Thus, the classifier was trained on more data from classes 1 and 5, compared to the other classes, and performed better at correctly identifying segments of those classes. Random selection

was chosen for this thesis to ensure that an initial bias would not be introduced to the ratings, as would have been the case had researchers manually selected segments belonging to each class to ensure class balance. Furthermore, the analysis reveals that there was much disagreement among raters regarding the classification of segments between classes 2-4, so any manual selection of data segments may not have actually yielded balanced classes regardless. Therefore, we opted to randomly select segments to be used for the gold standard.

While it can be argued that there are no formal experts to provide visual assessments of PPG data, the raters used in the establishment of the gold standard were not all familiar with PPG signals. Though they were all biomedical engineers, some had no prior experience analyzing PPG and were thus entirely reliant on the class examples provided. The effect of this was mitigated by establishing clear definitions for each quality class for the raters to follow.

The final classifier itself had low performance for data belonging to classes 2 and 3. Thus, there is greater uncertainty about data classified by the system into those two classes.

8.2 Future Work

The next phase of the project is to collect data from patients with cardiovascular conditions such as atrial fibrillation to determine whether the system can be used to distinguish between normal sinus rhythm and atrial fibrillation. This would be done in conjunction with the algorithm developed in this thesis, with the data classified to class 5 being used for arrhythmia analysis. Future work should also be done on restoring mixed quality data (classes 3 and 4) so that they may be usable for cardiovascular monitoring, as this would greatly improve the potential of the technology. If

successful, the system may be used in further medical studies to explore the progression of atrial fibrillation and other cardiovascular illnesses over the long-term, which could lead to the development of predictive systems. This would enable medical interventions to be applied earlier in the progression of the illness, thereby potentially preventing problems cardiovascular problems such as strokes and heart attacks.

Future work may also be done to expand the analysis done in Chapter 7. Area plots similar to Figure 7.1 may be generated to understand the variation in individual signal quality trends, comparison of individual "class 5 zones" (delineated by vertical black lines in Figure 7.1), and to identify outliers for each of the demographic groups. Further analysis can be done to explore the links between activity levels and signal quality throughout the 24-hour period. The data from the 3-axis accelerometer could be used as a representation of activity levels, with the per class percentages analyzed in terms of activity level. The inclusion of an activity log in future data collection would further enhance our understanding signal quality issues related to daily activities.

Appendix A

Random Forest Classifier

The final *Random Forest* classifier consists of 20 *Decision Trees*, operating in a plurality voting system. In each tree, there are three types of nodes: root, internal, and leaf. The root nodes have no parent nodes, as they are the starting points for each tree. The leaf nodes are the class assignments, used as termination points of the trees. The root and internal nodes use feature-based conditions to evaluate branching decisions. The pseudo-code for the nodes in each of the 20 *Decision Trees* used in the final classifier in Chapter 7 is available at `http://www.sce.carleton.ca/faculty/adler/publications/2016/npradhan-2016-MaSc-thesis-RFCode.txt`. A sample of the pseudo-code for Tree 1 is shown in Listing A.1. A partial visualization of Tree 1 is shown in Figure A.1.

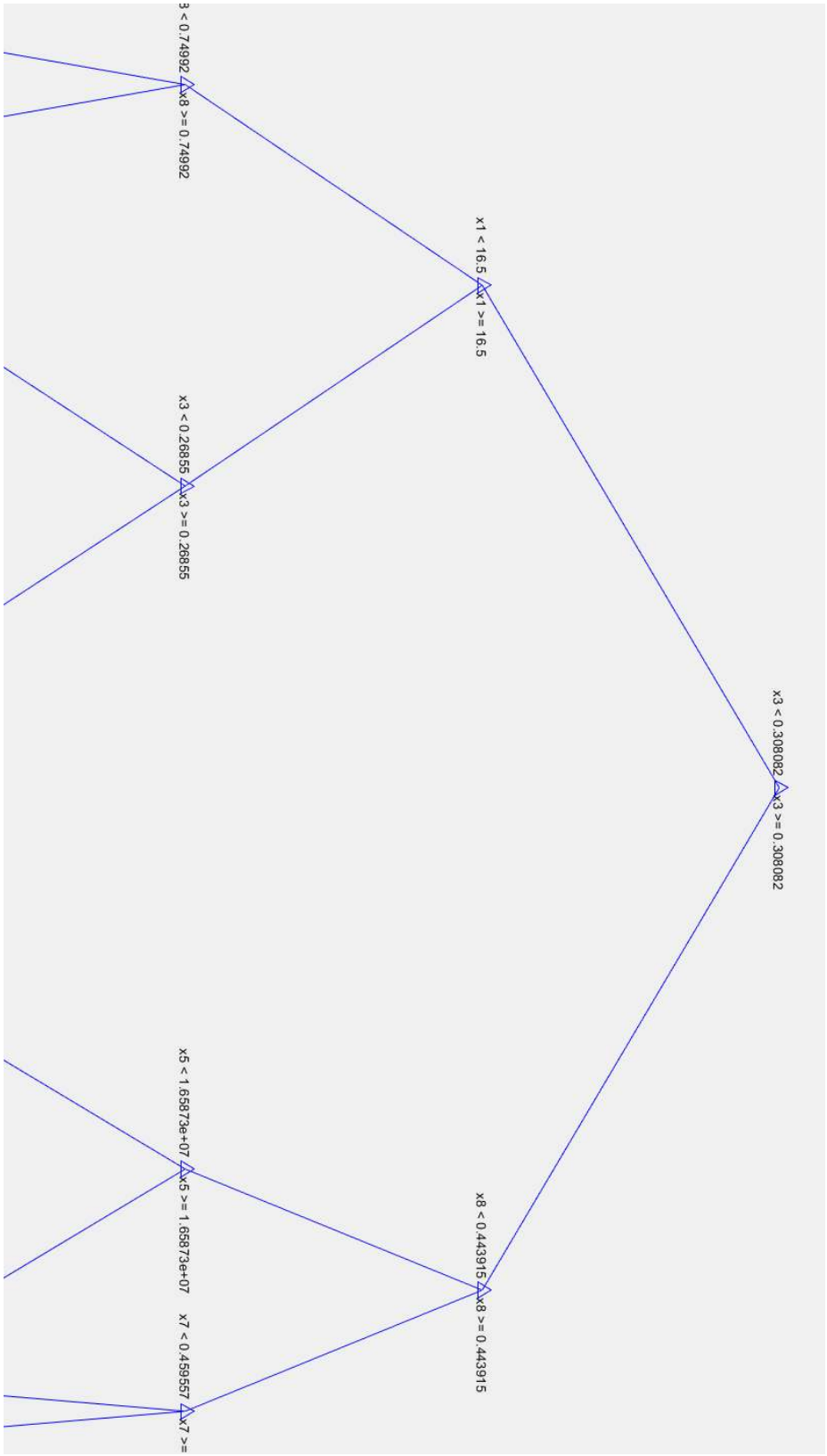

Figure A.1: Partial visualization of Tree 1 from the final *Random Forest* classifier.

```
  {Tree 1}
Node1: %Node Type: Root
  If (medianN < 0.30808), Then goto Node2, Else goto Node3,  EndIf
Node2: %Node Type: Internal,  Parent: Node1
  If (BillauerPeaks < 16.5), Then goto Node4, Else goto Node5,  EndIf
Node3: %Node Type: Internal,  Parent: Node1
  If (ACPeakVals1 < 0.44391), Then goto Node6, Else goto Node7,  EndIf
Node4: %Node Type: Internal,  Parent: Node2
  If (ACPeakVals1 < 0.74992), Then goto Node8, Else goto Node9,  EndIf
Node5: %Node Type: Internal,  Parent: Node2
  If (medianN < 0.26855), Then goto Node10, Else goto Node11,  EndIf
Node6: %Node Type: Internal,  Parent: Node3
  If (stdevE < 16587313.825), Then goto Node12, Else goto Node13,  EndIf
Node7: %Node Type: Internal,  Parent: Node3
  If (stdevACC < 0.45956), Then goto Node14, Else goto Node15,  EndIf
Node8: %Node Type: Internal,  Parent: Node4
  If (stdevACC < 0.42991), Then goto Node16, Else goto Node17,  EndIf
Node9: %Node Type: Internal,  Parent: Node4
  If (BillauerPeaks < 12.5), Then goto Node18, Else goto Node19,  EndIf
Node10: %Node Type: Internal,  Parent: Node5
  If (ACPeakVals2 < 0.38777), Then goto Node20, Else goto Node21,  EndIf
Node11: %Node Type: Internal,  Parent: Node5
  If (ACPeakVals1 < 0.35451), Then goto Node22, Else goto Node23,  EndIf
Node12: %Node Type: Internal,  Parent: Node6
  If (ACPeakVals1 < -0.31019), Then goto Node24, Else goto Node25,
   EndIf
Node13: %Node Type: Internal,  Parent: Node6
  If (ACPeakVals1 < 0.29266), Then goto Node26, Else goto Node27,  EndIf
```

Listing A.1: Partial pseudo-code for Tree 1 of the final *Random Forest* classifier.

References

[1] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, "Validation of the empatica e4 wristband," in *Student Conference (ISC), 2016 IEEE EMBS International*, pp. 1–4, IEEE, 2016.

[2] N. Pradhan, S. Rajan, A. Adler, and C. Redpath, "Classification of the quality of wristband-based photoplethysmography signals," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 269–274, IEEE, may 2017.

[3] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, p. R1, 2007.

[4] 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.

[5] O. Abdallah and A. Bolz, "Adaptive Filtering by Non-Invasive Vital Signals Monitoring and Diseases Diagnosis," in *Adaptive Filtering Applications*, ch. 7, InTech, jun 2011.

[6] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable Photoplethysmographic Sensors - Past and Present," *Electronics*, vol. 3, pp. 282–302, apr 2014.

[7] A. B. Hertzman, "The blood supply of various skin areas as estimated by the photoelectric plethysmograph," *American Journal of Physiology–Legacy Content*, vol. 124, no. 2, pp. 328–340, 1938.

[8] R. R. Anderson and J. A. Parrish, "The optics of human skin," *Journal of investigative Dermatology*, vol. 77, no. 1, pp. 13–19, 1981.

[9] Y. Maeda, M. Sekine, and T. Tamura, "The Advantages of Wearable Green Reflected Photoplethysmography," *Journal of Medical Systems*, vol. 35, pp. 829–834, oct 2011.

[10] W. Cui, L. Ostrander, and B. Lee, "In vivo reflectance of blood and tissue as a function of light wavelength," *IEEE Transactions on Biomedical Engineering*, vol. 37, pp. 632–639, jun 1990.

[11] P. Mohapatra, S. Preejith, and M. Sivaprakasam, "A novel sensor for wrist based optical heart rate monitor," in *Instrumentation and Measurement Technology Conference (I2MTC), 2017 IEEE International*, pp. 1–6, IEEE, 2017.

[12] M. Vinall, "An Update on the Epidemiology of Atrial Fibrillation," in *MD Conference Express*, vol. 15, pp. 2–3, SAGE Publications, 2015.

[13] T. M. Munger, L.-Q. Wu, and W. K. Shen, "Atrial fibrillation.," *Journal of Biomedical Research*, vol. 28, pp. 1–17, jan 2014.

[14] L. Ferranti and R. Laureanti, "Atrial Fibrillation detection in PPG signal recorded through a wristband device," Master's thesis, Politecnico di Milano, 2015.

[15] V. Markides and R. J. Schilling, "Atrial fibrillation: classification, pathophysiology, mechanisms and drug treatment," *Heart*, vol. 89, no. 8, pp. 939–943, 2003.

[16] P. Kirchhof, S. Benussi, D. Kotecha, A. Ahlsson, D. Atar, B. Casadei, M. Castella, H.-C. Diener, H. Heidbuchel, J. Hendriks, *et al.*, "2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS," *European Heart Journal*, vol. 37, no. 38, pp. 2893–2962, 2016.

[17] J. Lee, B. A. Reyes, D. D. McManus, O. Mathias, and K. H. Chon, "Atrial fibrillation detection using a smart phone," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1177–1180, IEEE, aug 2012.

[18] "AF Classification from a short Single Lead ECG Recording: The PhysioNet/-Computing in Cardiology Challenge 2017," *PhysioNet*, `https://physionet.org/challenge/2017/`.

[19] J. A. Sukor, S. J. Redmond, and N. H. Lovell, "Signal quality measures for pulse oximetry through waveform morphology analysis," *Physiological Measurement*, vol. 32, no. 3, p. 369, 2011.

[20] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Physiological Measurement*, vol. 33, pp. 1491–1501, aug 2012.

[21] J. Wander and D. Morris, "A combined segmenting and non-segmenting approach to signal quality estimation for ambulatory photoplethysmography," *Physiological Measurement*, vol. 35, no. 12, p. 2543, 2014.

[22] Empatica, *E4 Wristband User Manual*, 2015. Available at `https://empatica.app.box.com/v/E4-User-Manual`.

[23] Empatica Support, *Utilizing the PPG/BVP signal*. Available at `https://support.empatica.com/hc/en-us/articles/204954639-Utilizing-the-PPG-BVP-signal`.

[24] GE Healthcare, *SEER Light/ SEER Light Extend Compact digital recorders for ambulatory ECG testing*, 2005.

[25] GE Healthcare, *Operator's Manual: SEER Light Ambulatory Recorder/Controller*, 2008.

[26] J. L. Fleiss, "Measuring nominal scale agreement among many raters.," *Psychological Bulletin*, vol. 76, no. 5, p. 378, 1971.

[27] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, pp. 159–174, 1977.

[28] M. Elgendi, "Optimal Signal Quality Index for Photoplethysmogram Signals," *Bioengineering*, vol. 3, p. 21, Sep 2016.

[29] E. Billauer, "Peak detection using MATLAB (non-derivative local extremum, maximum, minimum)," `http://www.billauer.co.il/peakdet.html`.

[30] N. Chernov, "Ellipse Fit (Direct method)," *File Exchange - MATLAB Central*, `https://www.mathworks.com/matlabcentral/fileexchange/22684-ellipse-fit--direct-method`.

[31] A. W. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct Least Squares Fitting of Ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, 1996.

[32] E. W. Weisstein, "Ellipse," *MathWorld-A Wolfram Web Resource*, `http://mathworld.wolfram.com/Ellipse.html`.

[33] M. Abdelazez, M. Hozayn, G. S. K. Hanna, and A. D. C. Chan, "Gating of false identifications in electrocardiogram based biometric system," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 338–343, May 2017.

[34] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*. Springer Briefs in Statistics, Berlin: Springer-Verlag Berlin Heidelberg, 2013.

[35] A. Groth, "Singular Spectrum Analysis - Beginners guide," *File Exchange - MATLAB Central*, `https://www.mathworks.com/matlabcentral/fileexchange/58967-singular-spectrum-analysis-beginners-guide`.

[36] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2014.

[37] E. Frank, M. A. Hall, and I. H. Witten, *The WEKA Workbench Online Appendix for " Data Mining: Practical Machine Learning Tools and Techniques " Morgan Kaufmann, Fourth Edition, 2016*. Morgan Kaufmann, fourth edition ed., 2016.

[38] I. Guyon, A. Elisseeff, and A. M. De, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[39] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[40] P. Cunningham and S. J. Delany, "k-Nearest neighbour classifiers," *Multiple Classifier Systems*, vol. 34, pp. 1–17, 2007.

[41] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[42] C. Neuburger, "Multi Class SVM," *MATLAB Central - File Exchange*, `https://www.mathworks.com/matlabcentral/fileexchange/39352-multi-class-svm`.

[43] K. P. Murphy, "Naive Bayes classifiers," tech. rep., University of British Colombia, 2006.

[44] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. Pearson Education India, 2006.

[45] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," *IEEE Transactions on Geoscience Electronics*, vol. 15, pp. 142–147, jul 1977.

[46] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.