RANDOM FOREST CLASSIFIER FOR DETECTING NOISY EPOCHS IN RAW PPG SIGNAL FROM A SMARTWATCH

By Maryam Bayatzadeh

A thesis submitted to the Department of Computer Science
in conformity with the requirements for the degree of Master of Science
Bishop's University Canada
August 2023

Abstract

Photoplethysmography (PPG) is a low-cost, non-invasive technique used to detect volumetric changes in blood circulation, commonly used in smartwatch hardware for measuring heart rate. However, PPG signals are susceptible to motion artifacts, resulting in an unreliable heart rate estimation when the subject is moving and the sensor loses contact with skin. Many approaches have been proposed for resolving this problem, including the use of accelerometer data as a 'ground truth' for motion-corrupted epochs. Here, we propose a random forest classifier for detecting noisy epochs from the PPG signal alone. We first construct a dataset consisting of PPG signals collected from smartwatches worn by 22 participants over 30 days. We then constructlabels for the ground truth dataset based on the smartwatch accelerometer, which allows us to recognize epochs of high-motion activity. Finally, we trained the classifier to recognize noisy epochs based on the PPG signal alone (without using the accelerometer). Our best classification results yield 88.72% accuracy and 93.5% F1 score, showing that it is possible to recognize noisy epochs based solely on the attributes of the raw PPG signal.

Acknowledgments

I would like to thank my supervisor, Dr. Russell Butler, for his unfailing support, patience, and advice, without which none of this would have been possible. I would also like to convey my heartfelt gratitude to other professors in the Department of Computer Science at Bishop's University. I am particularly grateful to the distinguished faculty members of the computer science department and my defense committee for their insightful recommendations and ongoing assistance. Additionally, this endeavor would not have been possible without the support from the Bishop's Graduate Entrance Scholarship Foundation, which financed my research. Last but not least, my family deserves endless gratitude: my parents for teaching me to care about the right things as our lives and resources are finite, and I want to express my heartfelt gratitude to my spouse for his unfailing love, encouragement, and beliefin me, even when I doubted myself.

Table of Contents

List of Figures

List of Tables

List of Equations

Chapter 1: Introduction

1.1 Smartwatch technology

Smartwatches have emerged as game-changing wearable devices, elegantly combining timekeeping, biometrics, and powerful computing capabilities. This technology has its origins in the early 1970s, when Pulsar, a branch of Seiko, released the first digital watch. However, it wasn't until the 2010s that the concept of smartwatches fully took hold, with the release of devices such as the Pebble and the first Samsung Galaxy Gear. As of 2023, smartwatch use has skyrocketed, with millionsof individuals around the world utilizing these gadgets. The main reason for their appeal is convenience: smartwatch users can receive notifications, measure physical activity, monitor heart rate, and manage calls all from the convenience of their wrists. These characteristics make them useful tools for multitasking and leading a healthier lifestyle [1],[2].

Many firms have entered the smartwatch market, with big players like Apple, Samsung, Garmin, Fitbit, and Xiaomi at the forefront. Each organization has its own set of features, aesthetics, and capabilities that appeal to different consumer tastes. Design, distinctiveness, and screen size are major components of aesthetic appeal [3].

The photoplethysmogram (PPG) sensor, which detects blood flow through the skin, is a critical component of smartwatches. This sensor monitors heart rate and is essential for a variety of health-related applications. Smartwatches also include other sensors, such as accelerometers for activity tracking, GPS for location services, and gyroscopes for gesture detection [4].

The connection of smartwatches with smartphones, which allows users to access apps and transfer data smoothly, is further propelling their adoption. The future of smartwatch technology is bright, with continued developments and collaborations between industry titans constantly pushing the boundaries of what these wearable devices can achieve [5].

1.2 Heart rate, heart rate variability and disease

The major job of the heart is to circulate blood throughout the body, ensuring that tissues receive the oxygen and nourishment they require. Assessing heart rate (HR) and heart rate variability (HRV) is an important approach to determining heart health and general wellness. HR is the number of heart beats per minute, whereas HRV is the variability in interval between each heartbeat. HR and HRV are useful indications of the autonomic nerve system's (ANS) functioning, which governs many of the body's automatic functions such as breathing and digestion. HRV specifically measures the balance of the parasympathetic and sympathetic nerve systems – the 'rest and digest' and 'fight or flight' systems, respectively [6] [7].

A high HR and a low HRV constantly may indicate a lack of flexibility to stimuli, which is common in chronic diseases. Increased HR and lower HRV have been associated with an increased risk of cardiovascular disease (CVD), diabetes, and other chronic diseases. A high HR and low HRV in diabetic individuals suggest the existence of autonomic neuropathy, a condition in which the nerves that control the heart and blood vessels are destroyed, increasing the risk of CVD. This implies that HR and HRV could be utilized to detect early warning signals of cardiovascular disease in diabetics [8].

CVD, particularly heart disease, is also closely linked to increased HR and decreased HRV. These alterations could be attributed to the heart's decreased efficiency and overall poor health in people with these disorders. High resting heart rates, for example, have been linked to an increased risk of heart attack and stroke [9]. The link between HR and HRV in neurodegenerative illnesses such as Alzheimer's is complicated. Lower HRV may be an early sign of the disease, showing impaired function of the Vagus nerve, which is important in maintaining heart rhythm. However, more research is required to establish these findings [10].

While our understanding of these linkages is evolving, it is clear that the autonomic nervous system, as indicated in HR and HRV, plays an important role in chronic disorders. Recognizing this could lead to novel preventative tactics or therapies for persons suffering from or at risk of developing these illnesses. Using the PPG sensor, smartwatches provide a low-cost, non-invasive alternative to ECG for monitoring cardiovascular parameters like HR and HRV. However, PPG sensors on the wrist also pick up motion artifacts (MA), affecting the data's accuracy. Many methods to correct these artifacts use accelerometers to identify and ignore noisy data epochs during HR computation [11].

1.3 Goal of research

In this project, our goal is to devise a new approach to MA detection during smartwatch recording, based on the raw PPG signal and using machine learning. We first use the accelerometer signal to create a ground truth classifying PPG epochs as 'MA' or 'non-MA'. After creating our dataset, we generate features for training and testing to classify the data using the random forest.

1.4 Thesis outline

In this thesis, we first review the fundamental concepts of PPG signal and its relation to the accelerometer signal (Chapter 2). In Chapter 3 we present the dataset and methods of extracting smartwatch data features, then use those features for detecting motion artifacts in our PPG signal. Finally, in Chapter 4 we present the results of our classification methods and a short conclusion in Chapter 5.

Chapter 2: Background

2.1 Photoplethysmography

Photoplethysmography (PPG) is a non-invasive technique used to detect blood volume variations through an infrared light sensor placed on the surface of the skin. Correct identification of the PPG waveform and its main features is essential to extractseveral biomarkers, such as heart rate, blood pressure, cardiac output, and blood oxygen saturation. PPG sensors are usually placed on the distal parts of the human body, such as the arms, wrists, fingers, feet, or ears [12]. Some key benefits of using PPG in smartwatches are HR monitoring, sleep monitoring, stress, and relaxation monitoring. However, it is well known that motion artifacts (MA) can distort the signal. Some important noise sources that affect PPG analysis include body movement and sensor attachment, baseline change due to respiration, and hypoperfusion due to decreased peripheral perfusion.

Figure 1 shows the basic features obtained directly from the PPG waveform. The systolic peak in a PPG signal refers to the highest point or peak in the waveform during the systolic part of the heart cycle. The systolic phase is characterized by the contraction of the ventricles of the heart and the ejection of blood into the arteries. Systolic blood pressure is generally measured using the systolic peak, which indicatesthe highest amplitude of pulsatile blood flow. The dicrotic notch is a slight downwarddeflection in the PPG waveform that occurs shortly after the systolic peak. It is causedby the aortic valve closing, which temporarily blocks arterial blood flow. The dicrotic notch is generally visible as a secondary peak or a tiny dip in the PPG waveform's descending limb. It denotes the start of the diastolic phase and is a key factor in determining arterial flexibility and vascular resistance. The highest point or peak in the waveform during the diastolic phase of the cardiac cycle is referred to as the diastolic peak in a PPG signal. The diastolic phase is characterized by the relaxation of the ventricles of the heart and the filling of the coronary arteries. The diastolic peakis used to calculate diastolic blood pressure because it represents the largest amplitudeof blood flow during diastole [13].

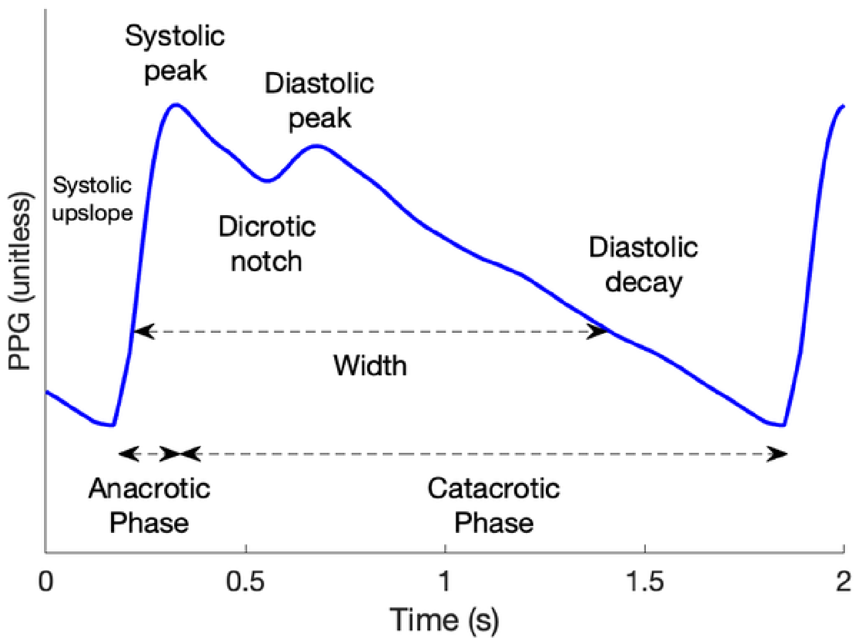

Figure 1: Major features of the pulse wave recorded using PPG. Figure reproduced from Charlton et al, 2021.

Any wearable devices acquire additional signals which can be used with the PPG to provide improved physiological monitoring. Figure 2 shows one of the additional signals commonly acquired by wearables. As we can see in Figure 2a and Figure 2b the accelerometer signal increases during times of high motion, and the corresponding PPG signal is contaminated by noise during those timepoints, making accurate HR estimation impossible

Accelerometry: Accelerometers measure static and dynamic acceleration. Due to their power efficiency and low price, they are already used in the majority of wearables for step counting and recognizing some activities (such as walking). Accelerometrysignals can also be used to improve parameter estimation from the PPG [14].

Accelerometry signals can be used to reduce noise in PPG signals by canceling noise which is common to both accelerometry and PPG signals. Accelerometry can also be used to identify periods when activity levels are too high to estimate parameters reliably from PPG signals. Accelerometry could also be used to contextualize PPG-derived parameters according to the activity in which they were recorded, as accelerometry can be used to infer body position (e.g. lying or standing), and could be used to identify a wide range of activities of daily living [15].

Gyroscope: Gyroscopes measure angular velocities around orthogonal axes and thus are suitable for capturing rotational movements. This feature can be used for adaptive motion artifact cancellation from PPG signals. In addition, gyrocardiography was recently proposed as a noninvasive monitoring method for the assessment of cardiacmechanics.

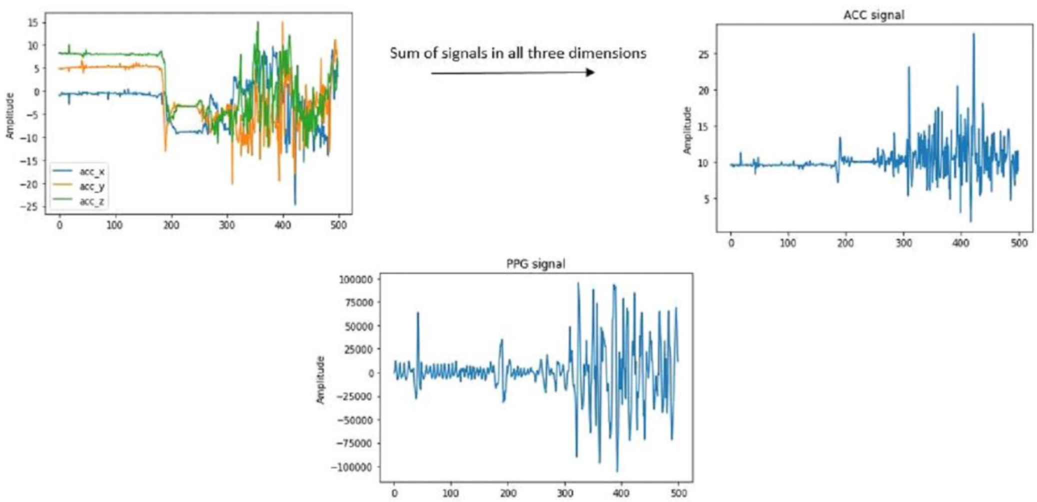

Figure 2: Motion artifacts. Top left: accelerometer signals in three dimensions (x ,y, z). Top right: sum of accelerometer signals. Bottom: PPG signal taken from same time period, showing motion artifacts from samples 300-500.

2.2 Types of noises in PPG signals

There are three types of noises in PPG signals shown in Figure 3. In this section, we explain each of them briefly.

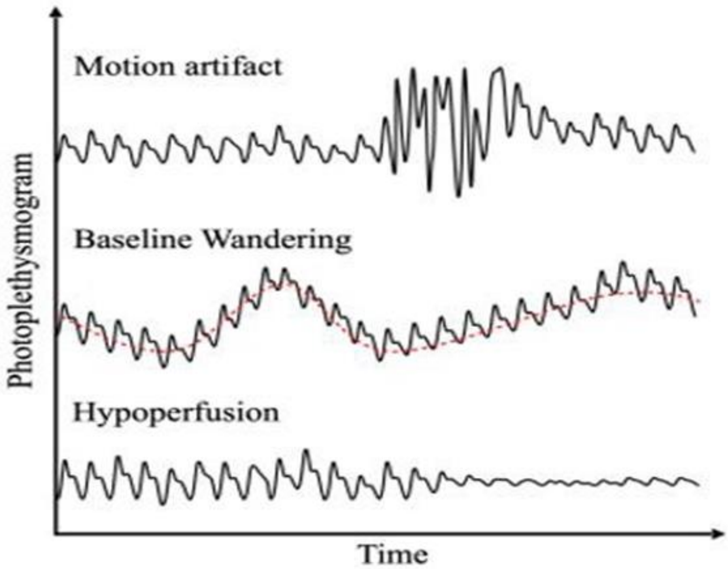

Figure 3: Examples of representative PPG distortion due to motion artifact, baseline wandering, and hypoperfusion. Figure reproduced from Shin et al, 2022.

2.2.1 Motion artifact

Motion artifacts, mainly caused by body motion such as hand movement, walking, and running, are critical noises when measuring PPG. When a person moves, motion artifacts are introduced into the PPG signal. These artifacts can be caused by mechanical movements of the skin, muscle contractions, or vibrations. Motion artifacts manifest as noise or interference in the PPG signal, making it difficult to extract accurate information about blood volume changes. Depending on probe type and light source, PPG measurement may be more sensitive to MA. Since MA is known to have a frequency range of 0.01–10 Hz, the major component of PPG can be distorted by overlapping with the main frequency band (0.5– 5 Hz) of PPG. Such distortion makes it difficult to detect important features during analysis, causing false diagnoses. Therefore, MA must be removed or corrected before analysis. In MA removal using a frequency domain filter, a high-pass filter (HPF) is mainly used.

2.2.2 Baseline and DC component

The baseline or DC component of the pulsatile component of PPG and the AC amplitude of PPG can be changed by various factors, such as respiration, sympathetic nervous system activities, and thermoregulation. The DC component is the constant or average value in a time-varying signal. It shifts the signal up or down but doesn't affect its shape. HPF is frequently performed in the method of directly removing the baseline. The frequency component of the AC of PPG is a component related to pulsation. This is normally higher than 0.5 Hz (30 bpm) in a healthy person. However, the respiratory component that causes baseline change has a frequency range of 0.15–0.5 Hz. HPF is performed to remove baseline movement located in the low-frequency range without damaging the AC component, based on the frequency range difference of signals. HPF is simpler and more convenient to perform than the method of baseline removal based on direct estimation. Hypovolemia, hypothermia, vasoconstriction, and decreased cardiac output or meanarterial pressure may weaken changes in blood volume in blood vessels, called poor perfusion or low perfusion.

2.2.3 Hypoperfusion

A medical disorder known as hypoperfusion is characterized by inadequate blood transport to tissues and organs, frequently leading to a shortage of the nutrients and oxygen required for cellular function. Tissue damage and, in extreme circumstances, organ failure may result from this. Hypoperfusion becomes more pronounced toward the peripheries of the body. It affects the pulsatile component of PPG, thus weakening amplitude change.

In addition to the movement, respiration, and low perfusion of a subject, numerous factors can distort the PPG waveform. Typical examples include ambient light, the temperature of the measuring site, skin pigmentation in the measurement of body size, alignment of the light source and photodetector, method of attaching the sensorto the skin, the contact pressure between the sensor and the skin, and posture of a subject. Ambient noise reduction is mainly attempted through hardware improvement [13]. Here, we focus mainly on detecting and remove the MA, which can be particularly challenging when computing derived featuresfrom the PPG waveform like HRV [16].

Chapter 3: Literature review

Nowadays, PPG signals can easily be collected continuously and remotely using inexpensive, convenient, and portable wearable devices (smartwatches, rings, etc.) which makes them a suitable source for wellness applications in everyday life. However, PPG signals collected from portable wearable devices in everyday settingsare often measured when a user is engaged with different kinds of activities and therefore are distorted by motion artifacts. There exists a variety of methods to detectand remove motion artifacts from PPG signals. The majority of the works related to the detection and filtering of motion artifacts in PPG signals can reside in three categories: (1) non-acceleration-based, (2) using synthetic reference data, and (3) using acceleration data. We explain each of them in this chapter.

3.1 Non-Acceleration-Based Methods

The non-acceleration-based methods do not require any extra accelerometer sensor for motion artifact detection and removal. In existing works, these approaches utilize certain statistical methods because some statistical parameters such as skewness and kurtosis will remain unchanged regardless of the presence of the noise. In [17], such statistical parameters are used to detect and remove the impure part of the signal due to motion artifacts. If there is movement, the amplitude of the PPG signal changes greatly, and consequently, the statistical parameters rise above the formerly set thresholds and the signal is marked as corrupt. The corrupted signal is then cut out of the original signal and only the clean signal is left, as can be seen in Figure 4. According to a set threshold, a part of the signal is marked as corrupted with movement and cut out. In [18], authors detect motion artifacts using a Variable Frequency Complex Demodulation (VFCDM) method. In this method, the PPG signal is normalized after applying a band-pass filter. Then, to detect motion artifacts, VFCDM distinguishes between the spectral characteristics of noise and clean signals. Then, due to a shift in the frequency, an unclean-marked signal is removed from the entire signal. Another method in this category is proposed in [19] that uses the Discrete Wavelet Transform (DWT) method. The DWT is used in signal processing to analyze several features (or 'frequencies') of the signal. This is accomplished by splitting the signal into lower-resolution (low-frequency) and higher-resolution (high-frequency) components. Each degree of decomposition corresponds to a different scale or frequency of information. The low-frequency components of the signal correspond to slow changes in the signal, such as the overall trend, and the high-frequency components correspond to 'rapid' changes, such as noise or tiny variations. The noise-free reconstructed PPG signal is produced by combining the low-frequency (trend) and high-frequency (variation) components. The 'mother wavelet' - the basic waveform utilized in the transformation process - is often picked before the signal is analyzed, which is a crucial limitation in adopting DWT. Because this waveform is predetermined, it may not fully match the PPG signal's unique characteristics, especially as the signal can be non-stationary (its statistical properties shift over time)and non-linear (it does not follow a straight line). As a result, some vital physiological data may be lost during the study.

In most non-accelerometer-based methods, the clean output signal is often shorter than the original signal, since unrecovered noisy data is removed from the signal.

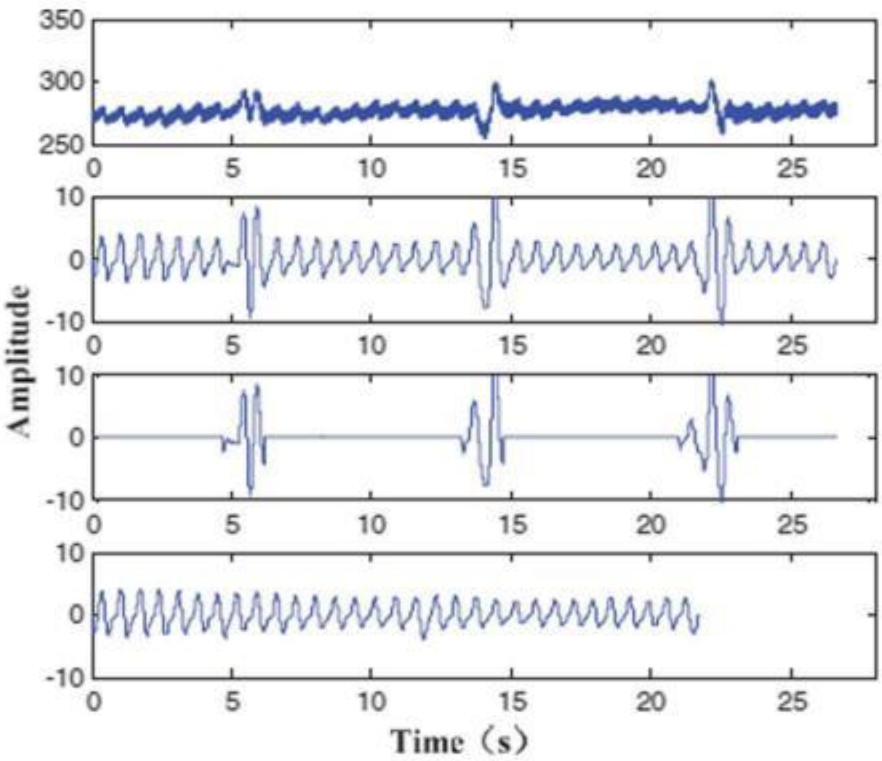

Figure 4: PPG preprocessing. Top: original PPG signal. Second row: PPG after pre-processing. Third row: detected movement. Bottom: after cut-out algorithm applied. Reproduced from TaheriNejad et al, 2019.

3.2 Synthetic reference data approaches for noise removal

The length of the clean signal with some non-accelerometer-based methods is shorter than the original signal. To mitigate this problem, a synthetic reference signal can be generated from the corrupted PPG signal. In [20], the authors use Complex Empirical Mode Decomposition (CEMD) to generate signals. After pre-processing, the EMD can be applied to the PPG signals. This entails breaking down the complicated PPG signals into a series of intrinsic mode functions (IMFs). As a parallel to simple harmonic functions, each IMF represents a simple oscillatory mode. This is accomplished by detecting local maxima and minima iteratively and then generating an envelope defined by a cubic spline line. The original signal is subtracted from the envelope mean result, and the procedure is repeated untilthe residual (the subtraction result) becomes a monotone function from which no more IMFs can be recovered. The EMD-derived IMFs can now be used as reference data. It's worth noting that not all IMFs are employed; rather, a choice is made based on frequency content or other heuristic reasoning. After obtaining the reference data,it can be utilized to decrease motion artifacts in the original PPG signal. This is often accomplished through the elimination of motion artifact-related IMFs, or the use of methods such as adaptive filtering or wavelet denoising, with the reference signal serving as a guide to the reduction of motion artifacts.

A similar approach that uses asynthesized reference signal was shown in [21]. A second PPG sensor is used to generate the movement signal. The second PPG sensor is positioned a few millimetersaway from the skin, as seen in Figure 5a, so it only measures when the subject is in motion. The second PPG sensor is intentionally located a few millimeters farther fromthe skin to capture primarily motion artifacts rather than the cardiac-related signal. This sensor does not detect any meaningful signal when the subject is immobile. Whenthe person moves, however, this sensor begins to take up signals due to its closeness to the skin, and these signals are primarily connected with motion artifacts. This second sensor's motion signal can then be used as a "reference" motion artifact signal.The signals collected from the two sensors can then be compared. The motion artifactscould then be statistically identified and eliminated from the primary signal, providing a cleaned PPG signal.

In Figure 5b one can see the typical outputs of both sensors, first few seconds withoutmotion and afterwards with motion.

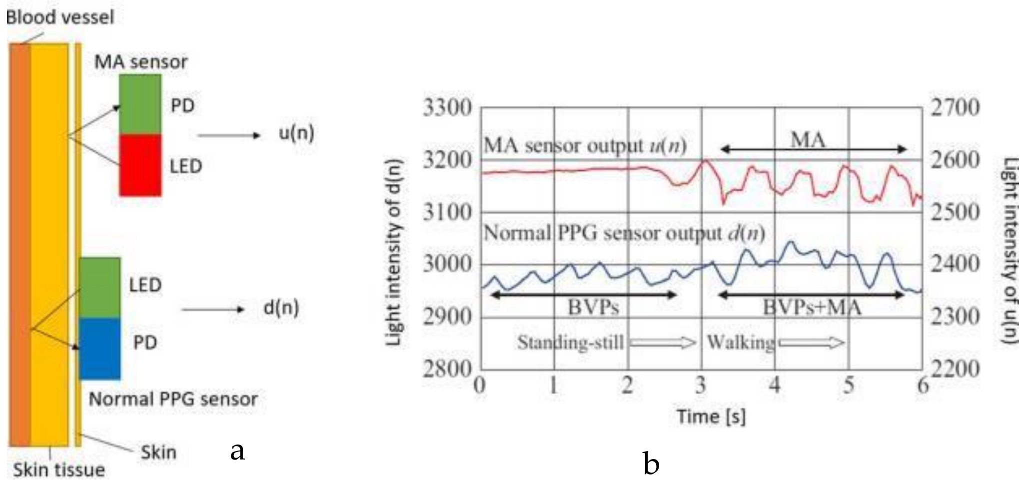

Figure 5: a) Application of the two PPG sensors on the skin, where the second sensor is a reference sensor used to generate the motion signal. Reproduced from TaheriNejad et al, 2019.

3.3 Acceleration-based methods

Often an accelerometer sensor is also embedded in wearable devices. To eliminate the effect of motion artifacts, acceleration data can be used as a reference signal. In [22], with the help of acceleration data, Singular Value Decomposition (SVD) is used for generating a reference signal for an adaptive filter. In the adaptive filtering process, SVD is used to provide a reference signal for motion artifact reduction. A correlation exists between the motion accelerometer data and the noisy components of the raw PPG signal. It is therefore critical to separate PPG band motion artifact components from accelerometer data. The motion artifact reference signal is generated using frequency components from the PPG signal frequency band. Least mean square (LMS) adaptive filters based on stochastic gradient descent are used to remove in-band motion artifacts. The noise reference for the filtering procedure is the intended reference signal created in the previous stage using SVD. The LMS method adjusts filter coefficients based on the current least mean error. Different LMS filter versions,such as Normalized LMS, Delayed LMS, and Adjoint LMS, are investigated to discover the best one for this task. Then, the reference signal and PPG signal pass through an adaptive filter to remove motion artifacts. Because of its simplicity and stable performance, the LMS adaptive filter is a popular type of adaptive filter. The LMS filter minimizes the mean square error between the desired and actual signals using a gradient-based steepest descent algorithm.

With a similar approach, authors in [23] use DC remover using another type of adaptive filter. Another method proposes a robust HR estimation technique, denotedas MURAD- multiple Reference adaptive noise cancellation for HR estimation when the PPG signal is severely corrupted by MA. First, MA is removed from the PPG signalusing an adaptive filtering technique based on recursive least squares (RLS). The key problem here is to develop an appropriate reference noise signal (RNS) for the adaptive filter. They employed 3-axis accelerometer data and the differential signal between the two PPG signals. Unlike previously reported adaptive noise cancellation(ANC) based algorithms (which employ a specific for a given time window), they made a real-time choice of an RNS for each time frame that is most likely to result in an accurate HR calculation. They were able to minimize the MA by using many RNSseven though the direction of motion changed from time to time, obtaining a separate version of the cleaned PPG signal for each RNS. They then estimate a set of plausible HR values for each time window using all of the cleaned PPG signals. Because HR is a slowly fluctuating signal with significant overlap between subsequent frames, the difference between the actual HRs of two consecutive windows is unlikely to be significant. As a result, the particular number from the probable HR set that is closestto the previous window's estimated HR is picked as the HR in the current window. The estimated HR in this method, combined with several heuristic peak verification techniques, may estimate the HR with excellent accuracy even when the participants conduct severe physical workouts such as running, leaping, boxing, swimming,weightlifting, and so on[24].

The other method used a combination of adaptive filters using a single noise reference (CASINOR), a collection of adaptive filters with a single noise reference signal, to estimate the HR from the corrupted PPG signal. The power value of the accelerometersignals is used to choose the noise signal (i.e., a single accelerometer signal) from the three accelerometer signals, Figure 6. Combining two adaptive filters, namely RLS adaptive filters, a slow filter, and NLMS adaptive filters, results in a rapid filter that increases the quality of the denoised signal. The RLS filter removes the MA component from the PPG signal, whereas the

NLMS filter enhances time-varying parameter tracking. The proposed denoising structure combines the RLS and NLMS adaptive filters, with a single accelerometer signal serving as the common reference noise signal. To remove the MA component from the preprocessed PPG data, the accelerometer signal with the highest power value is provided as a reference input tothe adaptive filters [21].

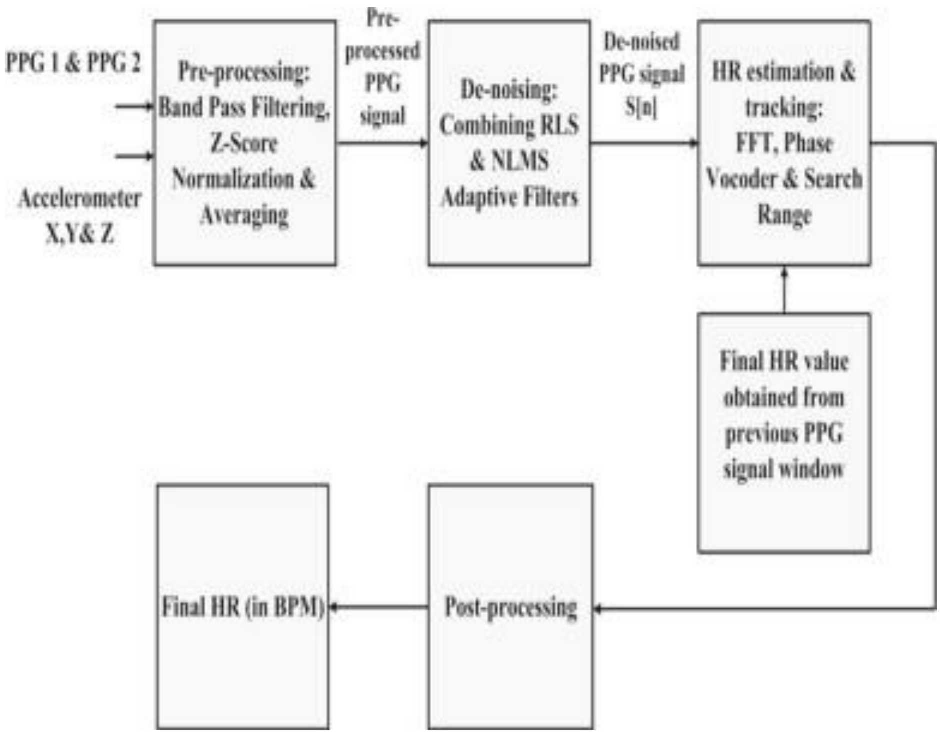

Figure 6: Block diagram of the proposed method used to estimate HR in [21]

In smartwatches like Apple, Garmin, and Fitbit, there are many ways to tackle the issue of noise and motion artifact cancellation, including: 1) Hardware Approaches: For example, Fitbit employs pure pulse technology, which entails beaming LED lightsonto the skin and measuring variations in light absorption with each heartbeat. This method can aid in noise reduction by ensuring a clear, robust signal in the first place. Also, The Apple Watch detects the quantity of blood flowing through the wrist at anyone time by pairing green LEDs with light-sensitive photodiodes. Apple Watch can compute the number of heartbeats each minute by flashing its LED lights hundreds oftimes per second. 2) Multi-sensor integration: Other sensors, like accelerometers and gyroscopes, are commonly found in Apple and Fitbit devices. The information gathered by these sensors can be utilized to increase the accuracy of heart rate measurements. For example, if the gadget detects that the wearer is moving a lot usingthe accelerometer, it can factor this into its analysis and perhaps reject PPG data collected during this period as it is likely to be noisy. 3) Software approaches: To remove high-frequency noise and smooth the data, signal processing and filtering techniques might be applied. Data that is likely to be noisy can be identified and excluded using algorithms. They may, for example, detect periods of intensive physical activity (using accelerometer data) and reject PPG data collected during theseperiods because it is likely to be influenced by motion artifacts.

As we mentioned previously, we want to use an accelerometer signal in a smartwatch(Samsung Galaxy Active 2) to label each segment of PPG signals and detect good(clean) and bad (motion artifact) segments. We then used those labels to train a classifier to recognize good and bad epochs based on raw PPG signal alone.

Chapter 4: Proposed Method

4.1 Data acquisition

In this project, we use the Samsung Galaxy Watch Active 2 (Figure 7) for recording data from subjects. The Samsung Galaxy Watch Active 2 is an advanced smartwatch, released in 2019, and allows access to raw PPG signals (unlike other smartwatches such as the Apple Watch series, which allows for heart rate extraction but does not give access to the raw signal). From the watch, we can extract the x,y,z components of the accelerometer, raw PPG signal, x,y,z components of the gyroscope, and corresponding timestamp of each sample. Samsung Galaxy Watch Active 2 relies on the Tizen operating system, a Linux-based mobile operating system developed and used primarily by Samsung. The 'TizenStudio' integrated development environment is used to develop apps for the watch, using the C programming language. The app developed has the following features:

- Record 1 minute of PPG data for every 10 minutes of elapsed time (~2.4 hours of data per day), not recording continuously to preserve battery life and avoiddevice overheating. The data was recorded with a 10Hz sampling rate.
- Save raw PPG data and timestamps to the file system of the watch in .csv file format.

Figure 7: Samsung Galaxy Active 2 watch

Figures 8 and 9 show two representative epochs (10 seconds each) recorded from daily activity from one of the subjects. In Figure 8, the accelerometer has large fluctuations and the corresponding PPG signal is noisy. In figure 9, the accelerometer has very small fluctuations and the corresponding PPG signal is clean.

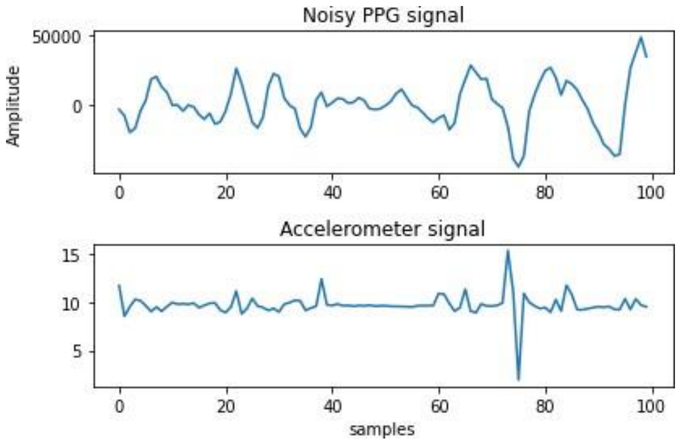

Figure 8: HRV=380ms and BPM=52.74, sampling frequency=10 Hz

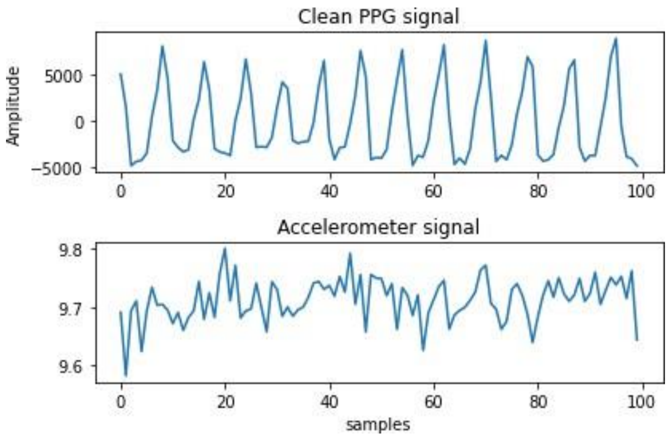

Figure 9: HRV=51ms and BPM=75.86, sampling frequency=10 Hz

The normal range of HRV for a healthy person is between 20 and 100 milliseconds. Also, the normal range of bpm for healthy adults is between 60 and 90 beats per minute. We calculate the HR (in BPM) and HRV (in ms) using the following method:

a) Calculating Heart rate

Peak Detection Method: Detect the peaks of the PPG waveform. The time between peaks is the period (T) of one cardiac cycle.

$$BPM = \frac{60}{T} \qquad (\;)$$

Where T is the time between peaks in seconds.

b) Calculating the HRV

SDNN (Standard Deviation of NN intervals): This is the standard deviation of the time intervals between adjacent peaks (often called NN intervals).

$$SDNN = \sqrt{\frac{\sum_{i=1}^{N}\left(NNi - mean(NN)\right)^2}{N}} \qquad (\;)$$

Where NNi is each NN interval and N is the total number of NN intervals.

4.2 Method for Labelling Dataset

The suggested technique labels motion artifacts in PPG signals based on accelerometer data. The accelerometer-captured motion of the hand serves as a reliable predictor of motion artifacts in the PPG signal. Accelerometers generate time-series data that represent the magnitude and direction of movement, which can be related to noise and artifacts in PPG data. As a result, accelerometers, which are typically included in the same wearable devices as PPG sensors, can be used successfully for this purpose.

Our method requires identifying 10-second data segments as 'epochs' and determining the standard deviation (STD) of accelerometer measurements for each epoch. Choosing a 10-second window for segmenting PPG signals was a balance between capturing enough data for robust motion artifact detection and allowing for real-time processing. Longer intervals could compromise the system's responsiveness, while shorter intervals might not provide sufficient data to accurately identify artifacts. Additionally, preliminary testing indicated that this segmentation duration yielded a reliable performance for artifact detection in our specific application. The standard deviation reflects the dispersion of data, and a greater STD indicates more intense or variable movement in the case of accelerometer data. As a result, epochs with more substantial hand movements, which are likely to cause moremajor motion artifacts, will have a higher standard.

Following that, we apply Z-score normalization to the obtained standard values. The Z-score indicates how far an element deviates from the mean. As a result, greater Z-score epochs correspond to periods of vigorous or irregular movement, which contributes to a larger degree of motion artifacts in the PPG signal. We can accuratelydesignate epochs with severe motion artifacts by evaluating Z-scores over a certain threshold. In this data, the threshold was $Z>2$. To locate outliers or abnormal data points, statistical data analysis frequently uses a Z-score threshold of 2. A Z-score of 2 is a cautious but useful cutoff for highlighting major deviations that may signify noise or motion artifacts because it corresponds to approximately 95.4% of the data lying within two standard deviations from the mean in a normal distribution. The decision is in line with the empirical rule in statistics, which is frequently applied to find outliers.

Using this approach, we can label all of the PPG segments that have motion artifacts by using an accelerometer. We then separate into our train and test dataset, and can classify each epoch of the PPG signal by using different machine learning algorithms.

4.3 Methods of classification

After creating the labeled dataset based on the accelerometer we used different methods for using this data for classification. By experiencing this method, we can find a fit mapping to the labeling method with an accelerometer signal. We used 1) raw data 2) FFT transformed data 3) Feature-selected data for doing the classificationand used different algorithms to find the best results for each of them. We discusseach of them below.

4.3.1 Extracting features related to the PPG signal

We extracted the following features from the raw PPG signal: mean, standarddeviation, skewness, kurtosis, and Heart rate variability (HRV):

a) Mean: The mean value of a signal is the average value of the signal:

$$Mean = \frac{\sum_{i=1}^{N} xi}{N} \qquad (\ )$$

The index of I is equal to index of the summation and N is the total number of elements you are summing. x_i represents each individual element (data point) that you're summing up.x_1,x_2,…,x_N are the elements you sum to calculate the mean.

b) Standard deviation: The standard deviation of a set of numbers is a measure of their variance or dispersion. A low standard deviation implies that the values of the set tend to be close to the mean (also known as the expected value), whereas a high standard deviation shows that the values are spread out over a larger range.

$$STD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(xi - \bar{x})^2} \qquad (\ )$$

Where $\bar{x}$ is the mean value of x.

c) Skewness: Skewness is a statistical term that refers to the asymmetry of a set ofdata values. It is used to calculate the extent and direction of skew (difference from horizontal symmetry) in data. The skewness of data distribution is zero ifit is fully symmetrical. A positive skew means that the right side of thedistribution's tail is longer or fatter than the left side's tail. In other words, the number of higher values is bigger. A negative skew, on the other hand, showsthat the tail on the left side is longer or fatter than the tail on the right side, indicating a greater number of lower values.

$$Skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(xi - \bar{x})^3}{std^3} \qquad (\ )$$

d) Kurtosis: Kurtosis is a statistical term that describes how data is distributed around the mean. It denotes the degree of 'tailedness' or the sharpness of a distribution's peak.

A distribution with a high kurtosis has a clear peak near the mean, a quick drop,and heavy tails. This is known as leptokurtic. A distribution with low kurtosis,on the other hand, has a flat top near the mean rather than a high peak. A Platykurtic distribution is flat. The standard normal distribution has a kurtosisof three and is known as mesokurtic.

$$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(xi - \bar{x})^4}{std^4} \qquad (\ )$$

e) Heart Rate Variability (HRV): The period between each heartbeat is used to calculate HRV. These intervals are commonly known as RR intervals. These intervals are commonly referred to as "pulse to pulse" intervalsor PP intervals in the context of PPG, but the concept is the same. We followed these general steps to compute HRV from a PPG signal:

- Identify each PPG signal peak that represents a heartbeat.
- Time the intervals between each peak (PP intervals). This can be done in avariety of units, but it is often done in milliseconds.
- Use the PP intervals to compute HRV.

In the first experiment, we extracted features for training and testing datasets. Then we trained the best model for finding the bad and good labeled data. Our goal is to find the best Model for the detection of bad and good epochs, thereby eliminating faulty HR estimates. We used an Extra trees Classifier for detecting the good and bad epochs in the dataset, details are below. We first explain the Random Forest Classifier.

4.3.2 Random Forest classifier

A random forest (RF) classifier is an ensemble classifier that generates several decisiontrees from a random subset of training samples and variables [25]. Before we can comprehend how the random forest algorithm works in machine learning, we must first grasp the ensemble learning technique. Ensemble simply refers to the combination of numerous models. As a result, rather than a single model, a group of models is utilized to create predictions. Ensemble employs two methods:

- Bagging: It generates a different training subset from the sample training data with replacement, and the final output is determined by majority voting. Consider Random Forest.
- Boosting: It turns weak learners into strong learners by constructing sequential models with maximum accuracy. For instance, ADA BOOST and XG BOOST.

Random forest operates on the bagging principle (discussed below).

4.3.3 Bagging

Random forest employs the ensemble approach of bagging, also known as Bootstrap Aggregation. Bagging selects a subset of the data at random. As a result, each model is constructed using the samples (Bootstrap Samples) provided by the original data using row sampling. This stage of row sampling with replacement is referred to as the bootstrap. Each model is now trained individually. After merging the findings of all models, the outcome is based on majority voting. Aggregation is the process of integrating all the findings and producing output based on majority voting (Figure10).

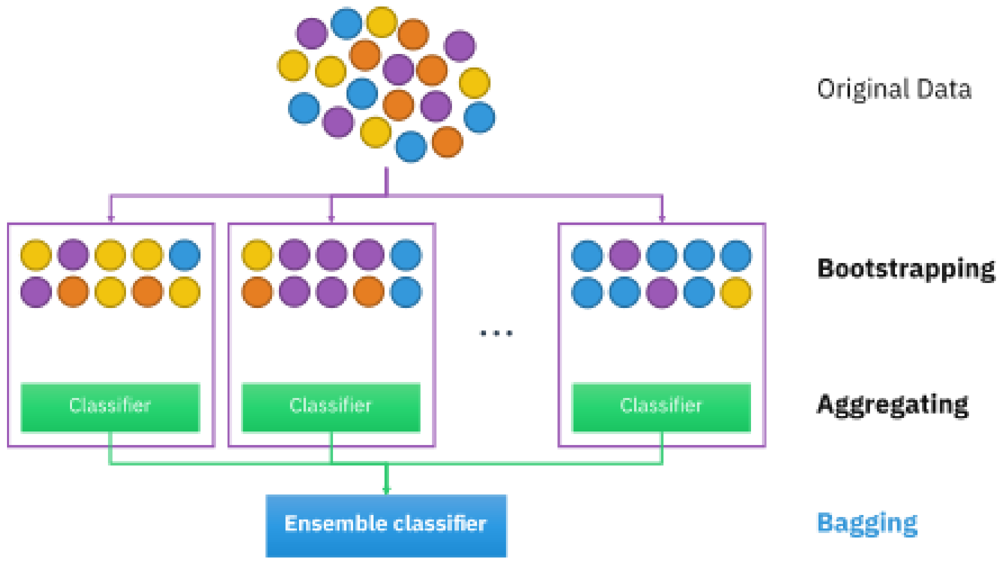

Figure 10: Bagging process. Reproduced from[31]

Figure 11 shows an example. The Bootstrap sample is taken from genuine data (Bootstrap sample 01, Bootstrap sample 02, and Bootstrap sample 03) with a replacement, implying that each sample may not contain unique data. This bootstrapsample's models (Model 01, Model 02, and Model 03) are trained individually. Contrasted with the sad emoji, the happy emoji now holds a majority. As a result of the majority vote, the final output is the happy emoji.

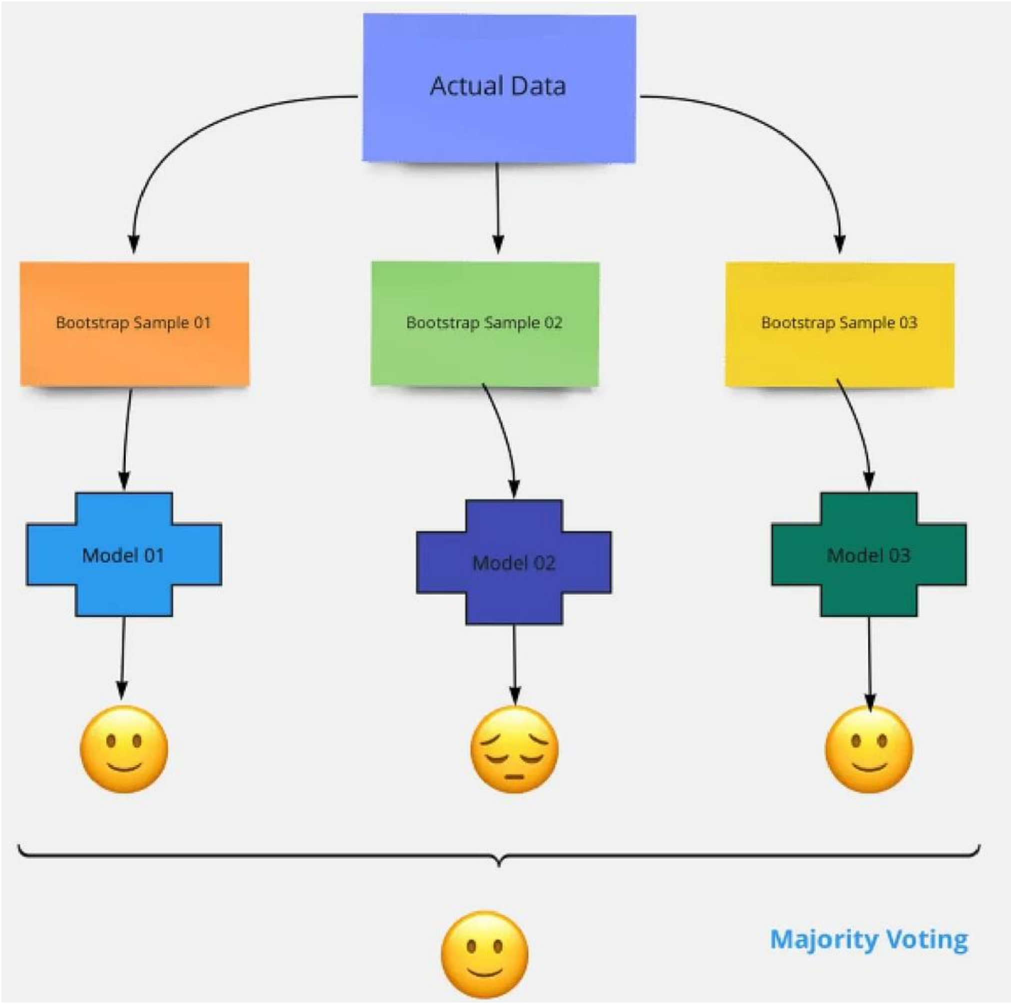

Figure 11: Bagging ensemble method. Reproduced from [32]

4.3.4 Boosting

Boosting is a strategy that employs the concept of ensemble learning. To obtain the final result, a boosting technique combines numerous simple models (also known as weak learners or base estimators). It is accomplished by developing a model in seriesutilizing weak models. There are various boosting algorithms; adaptive boosting (AdaBoost) was the first truly successful boosting algorithm designed for binary classification. AdaBoost is a popular boosting approach that combines numerous "weak classifiers" into a single "strong classifier", Figure 12.

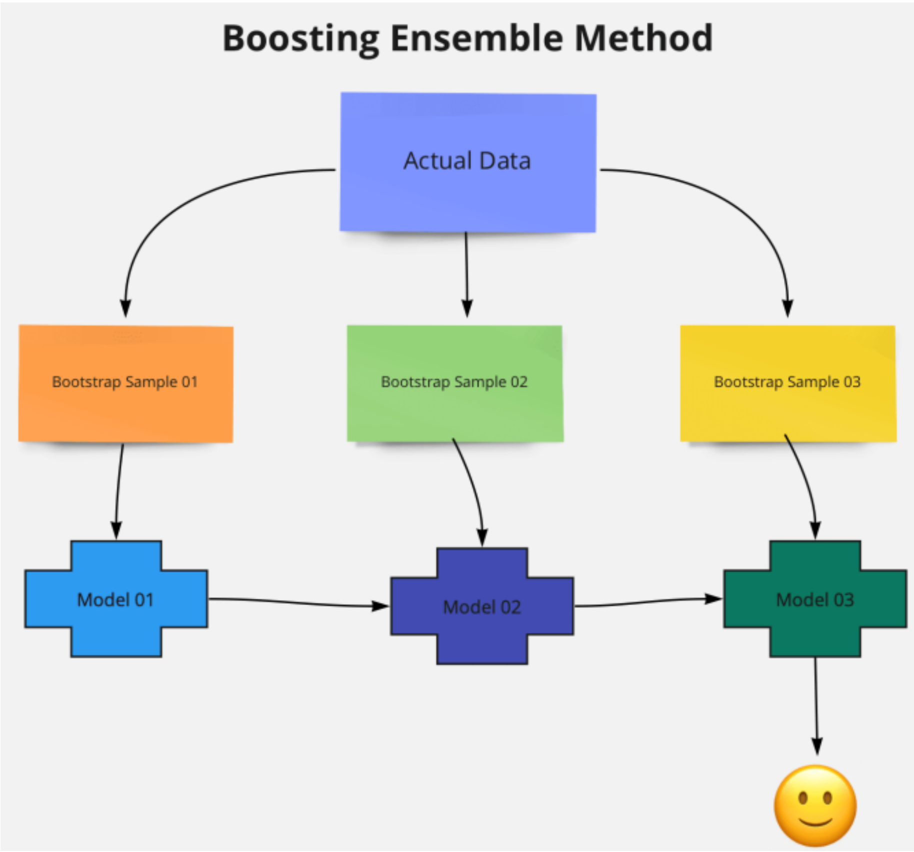

Figure 12: Boosting process. Reproduced from [33]

4.3.5 Random forest algorithm procedures

Step 1: For each decision tree in the Random forest model, a subset of data points anda subset of characteristics are chosen. Simply put, n random records and m features are drawn from a data set of k records.

Step 2: For each sample, an individual decision tree is built.

Step 3: Each decision tree will produce a result.

Step 4: For Classification and Regression, the final result is evaluated using Majority Voting or Averaging.

Consider the fruit basket as an example, as seen in the figure below. Now, n samples are drawn from the fruit basket, and an individual decision tree is built for each one. As illustrated in the illustration, each decision tree will produce an output. The final result is determined by majority voting. As seen in the graphic below, the majority decision tree produces an apple rather than a banana, hence the end outcome is an apple, Figure 13.

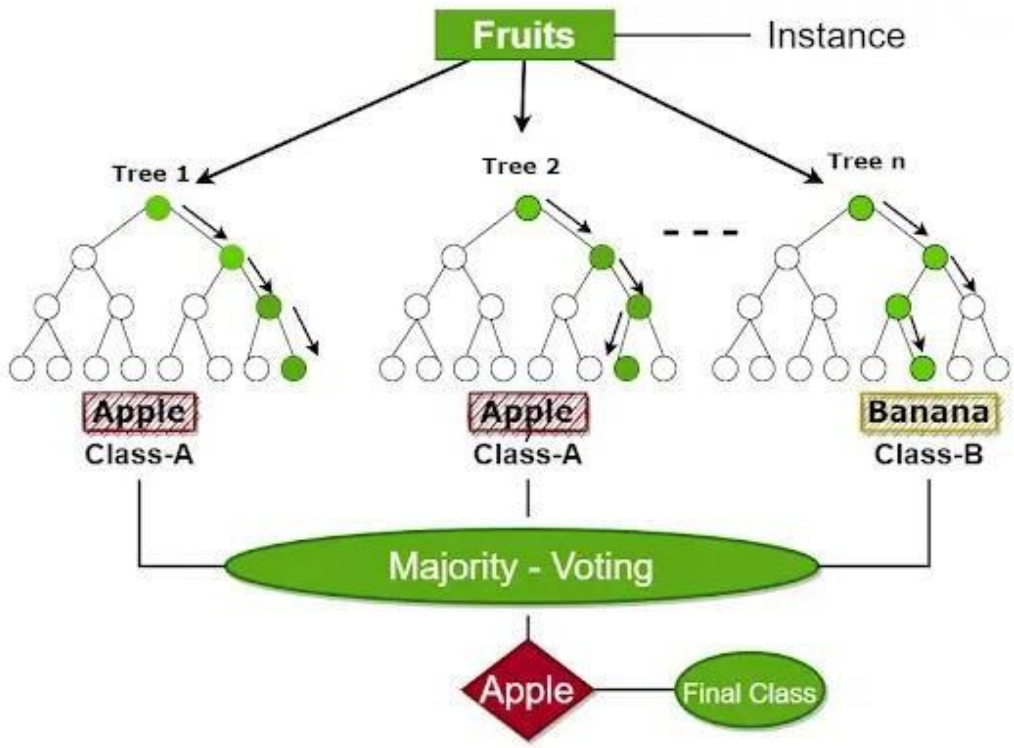

Figure 13: Example of Using Random Forest. Reproduced from[34]

4.3.6 Extremely randomized trees classifier

Similar to Random Forests, Extra Trees is an ensemble ML approach that trains numerous decision trees and aggregates the results from the group of decision trees to output a prediction. However, there are a few differences between Extra Trees andRandom Forest. Random Forest uses bagging to select different variations of the training data to ensure decision trees are sufficiently different. However, Extra Trees uses the entire dataset to train decision trees. As such, to ensure sufficient differences between individual decision trees, it randomly selects the values at which to split a feature and create child nodes. In contrast, in a Random Forest, we use a greedy algorithm and select the value at which to split a feature. Apart from these two differences, Random Forest and Extra Trees are largely the same.

The Extremely Randomized Trees classifier (commonly abbreviated as Extra Trees) isa form of ensemble learning algorithm that aggregates the outcomes of several de- correlated decision trees gathered in a "forest" into a single result. This method is similar to the more well-known Random Forest algorithm, but with one crucial difference: whereas Random Forest employs bootstrapping to find the best thresholdfor each randomly selected feature at each node, the Extra Trees classifier chooses these thresholds entirely at random.

This extra randomness can occasionally help to minimize model variance, which is useful if the model is overfitting. It can, however, increase bias, which might be problematic if the model is Underfitting. As a result, whether Extra Trees outperforms Random Forest is dependent on the specific data.

So what effect do these changes have? Using the entire dataset (which is the default setting and can be changed) allows Extra Trees to reduce the bias of the model. However, the randomization of the feature value at which to split increases the bias and variance. [14] introduced the Extra Trees model conducts a bias-variance analysisof different tree-based models. From the paper, we see on most classification and regression tasks Extra Trees have higher bias and lower variance than Random Forest.However, the paper goes on to say this is because the randomization in extra trees works to include irrelevant features in the model. As such, when irrelevant features were excluded, say via a feature selection pre-modeling step, Extra Trees get a biasedscore similar to that of Random Forest.

In terms of computational cost, Extra Trees is much faster than Random Forest. This is because Extra Trees randomly selects the value at which to split features, instead of the greedy algorithm used in Random Forest. [14] [26]

The second method we used for finding the better algorithm for classification isusing FFT which we will discuss about it below.

4.3.7 Using FFT of the dataset for detecting noise in signals

The other method that we tested was using a fast Fourier transform of training and testing datasets to obtain better classification results. Fast Fourier transform (FFT) is one of the most useful tools and is widely used in signal processing. The FFT is an algorithm for computing the Discrete Fourier Transform (DFT) of a digital sequence quickly. The DFT is a mathematical transformation that allows us to express a time- domain signal as a sum of frequency-domain sinusoidal functions. The DFT is a mathematical transform that separates the frequencies in a time-domain signal. FFT reduces the computational cost of DFT from $O(n^2)$ to $O(nlogn)$ allowing it to run much quicker on huge datasets. It decomposes a signal into its constituent frequencies,as well as their amplitudes and phases. FFT can be used to detect noise in the frequency domain. Noise is frequently manifested as sudden spikes at frequencies thatdo not correlate to physiological signals, or as a broad, continuous spectrum unrelatedto any specific biological function. In FFT, the following is the difference between a clean and a noisy signal: The FFT of a clean PPG signal should reveal clear and distinctpeaks corresponding to the heart rate and its harmonics. There may be additional peaks in a noisy signal corresponding to the noise sources, or the desirable peaks may be warped or broadened. If the noise is unpredictable, it might raise the baseline of the entire spectrum, making identifying the genuine signal more difficult [30]. The Fast Fourier Transform (FFT) mathematical formula:

If x[n] is an input vector of length N (n = 0, 1..., N-1), then the FFT is calculated asfollows:

For n = 0 to N-1, and k = 0 to N-1

$$X[k] = \sum_{n=0}^{N-1} x[n] * e^{\frac{j*2\pi kn}{N}} \qquad (\ )$$

In this formula, X[k] is the Fourier Transform value at frequency k (k = 0, 1..., N-1), and X[n] is the input signal value at time n (n = 0, 1..., N-1). In addition, 'j' represents the imaginary unit in complex numbers. For example, if we have a sequence x = [x[0], x[1], x[2], x[3]], we can compute its Fast Fourier Transform using the preceding method. Cooley and Tukey first proposed the concept of FFT in their seminal 1965 paper "An Algorithm for the Machine Calculation of Complex Fourier Series". This paper established the groundwork for modern FFT algorithms [27].

To summarize, FFT is an effective tool for analyzing PPG signals, and understandinghow noise expresses itself in the frequency domain can help guide noise reduction approaches. The FFT of a noisy signal will differ from that of a clean signal, reflectingthe presence and nature of the noise, and adequate analysis and preprocessing can aid in minimizing these undesired components.

After using FFT we found the best algorithm for the classification of the dataset and we found the Extra trees classifier that was explained in the previous section.

4.3.8 Raw data approach using Gradient boosting classifier

In this section, we used the raw data segments of signals for classification and then found the best algorithm for this. The algorithm we used for this section is the Gradient boosting classifier which is an ensemble algorithm.

The Gradient Boosting Classifier gradually constructs an ensemble of weak decision tree classifiers. It begins by developing a simple decision tree. Then it constructs successive trees in such a way that the previous trees' flaws are corrected. The new trees are fitted to the negative gradient of the previous predictions' loss function (usually the log loss or deviation). The Gradient Boosting Classifier's core componentsand concepts are as follows:

Boosting: As we mentioned earlier, "Boosting" refers to the approach of merging numerous weak learners (individual classifiers with somewhat better accuracy than random guessing) to build a strong classifier. The boosting technique increases performance iteratively by providing more weight to misclassified occurrences in eachiteration.

Gradient Descent: The term "gradient" in the phrase "Gradient Boosting" alludes to the gradient descent optimization technique used to minimize the loss function. The algorithm calculates the negative gradient of the loss function concerning the current ensemble's predictions in each iteration and fits a new tree to the negative gradient residuals.

Weak Learners: In the context of the Gradient Boosting Classifier, weak learners are often shallow decision trees with a short depth (also known as decision stumps), making them simple and computationally efficient. Based on a single characteristic and threshold, these trees make binary decisions.

Ensemble Learning: The Gradient Boosting Classifier's final prediction is a weighted combination (voting) of predictions provided by all of the individual decision trees inthe ensemble.

Regularization: To prevent overfitting and increase generalization, the Gradient Boosting Classifier can use regularization approaches such as restricting the depth of individual trees and employing a learning rate.

Because of its high predictive power and resistance to overfitting, the GradientBoosting Classifier is commonly employed in practice for classification tasks. It can handle both numerical and categorical features and works well on large datasets.

Chapter 5: Results

In this chapter, we present the results derived from the methods which arediscussed in the previous chapter. we divided them into three sections.

For the first experiment, we used 1500 segments which were labeled by using an accelerometer signal as noisy or clean signals. We used 22 subjects' samples whose data was recorded for one month and added these samples to the primary dataset to have robust results. After recording we applied a Band Pass filter to PPG signals between 0.5 and 5 Hz. After that, we used the method mentioned in the previous chapter (standard deviation of accelerometer) forlabeling every 10 seconds of PPG data. Every 10 seconds of PPG data includes 100 samples. For coding purposes, the label class "true" was assigned a value of 1 while the class "false" was replaced by a value of 0. After labeling the PPG segments, we used the three aforementioned methods (features, FFT, and raw) for training and testing.

During the preliminary examination of the experimental data acquired for this investigation, it became clear that there was a significant imbalance in the distribution of the two outcome labels [28]. This is a regular occurrence when using real-world of the classes.

To address this class imbalance, two basic methods were used: class weights and resampling procedures. The former entailed modifying the weights assigned to each class in the machine learning model so that the minority class was given more weight during the learning process. This strategy is intended to keep the model from being unduly influenced by the majority class. The resampling approach used, on the other hand, was the Synthetic Minority Over-Sampling Technique, or SMOTE, which is an oversampling technique that creates synthetic samples from the minority class ratherthan simply replicating existing examples. This method encourages greater diversity in minority class data and is intended to improve the model's performance on the underrepresented class.

A comparison of the resulting models revealed an intriguing observation following the comprehensive application of both class-weight and SMOTE approaches. The accuracy, precision, recall, and F1 scores of the models were comparable, showing that, for this specific dataset and situation, both techniques were equally effective in addressing the issue of class imbalance. These findings emphasize the need for robust strategies in minimizing class imbalance in machine learning initiatives thatuse real-world data.

5.1 Feature extraction results

After using the Extra trees classifier for this dataset after feature extracting we obtained **87.72%** accuracy and **93.26%**F1 score. In the imbalanced datasets, the F1 score is morereliable for classification. Also, the other measure for checking the reliability of the models is using the ROC curve. (The AUC-ROC curve is a performance metric for classification tasks at various threshold levels). AUC is an abbreviation for "Area Under the receiver operating curve (ROC).

The ROC curve is a graphical representation of a binary classifier system's diagnostic capabilities as its discrimination threshold is modified. It is calculated by graphing theTrue Positive Rate (TPR) vs. the False Positive Rate (FPR) at different threshold values. In machine learning, the True Positive Rate (also known as sensitivity, recall, or probability of detection) is the fraction of genuine positives that are properly detected. The False Positive Rate (also known as fall-out) quantifies the proportion of true negatives that are misidentified.) It is shown in Figure 14.

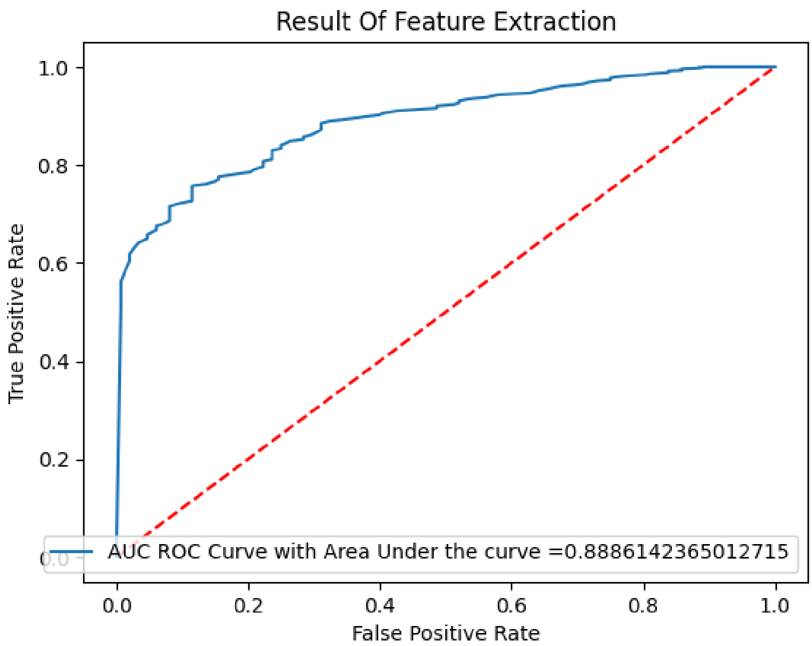

Figure 14: ROC curve for using feature extraction in the dataset.

In Tables 5-1 and 5-2 we show the features in the three good and bad labeled samplesin the dataset. these epochs are correctly classified by our model. As we can see in thetables, the measures are quite different for good and bad epochs. The statistical parameter differences show that bad epochs that have motion artifacts have a higher amount of parameters than epochs that don't have motion artifacts. Skewness and kurtosis are measures of a signal's asymmetry and tail extremities, respectively. The distribution can be distorted by noise and motion artifacts, resultingin increased skewness or kurtosis values.

Table 1: Features extracted for good epochs

Mean	Standard deviation	Skewness	Kurtosis	HRV
164.44	7119.2	0.48	1.17	57ms
254.82	5541.95	0.54	1.27	59ms
149.06	7542.01	0.49	1.06	53ms

Table 2: Features extracted for bad epochs

Mean	Standard deviation	Skewness	Kurtosis	HRV
750.17	135671.86	4.55	34.61	300ms
1919.87	83375.72	4.3	31.04	389ms
492.12	100514.17	2.88	21.9	269ms

We then found the most essential feature that has the most effect on training thedataset (Figure 15). HRV had the most effective role in the training of data.

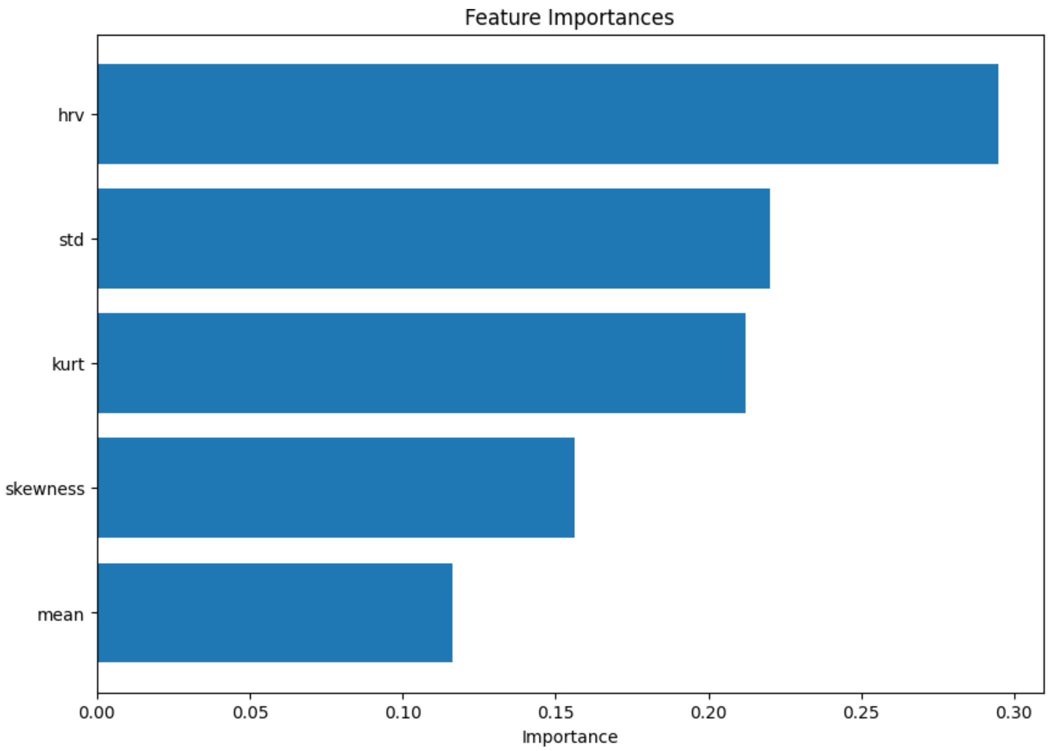

Figure 15: Comparison of feature importance

5.2 Using FFT of the dataset

The second method we tested was running the classifier on the FFT of our epochs.

We used this transformation on our training and testing dataset to see the results. wegained **86.4%** accuracy and **92.5%** F1 by using FFT for classification. As we can see inthe figure in the roc curve the AUC (area under the curve) is about **88%** which is a good measure for this classification (Figure 16).

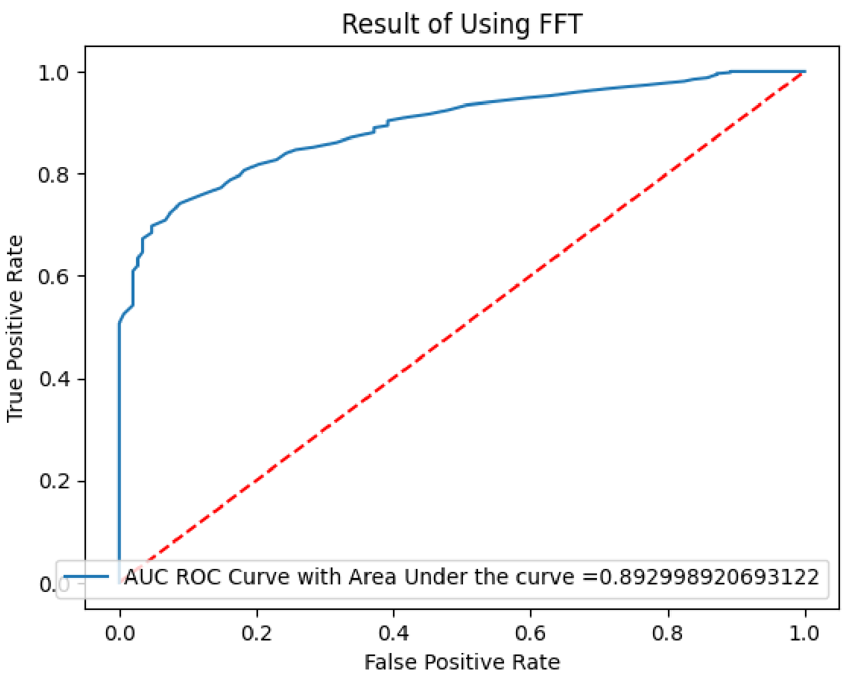

Figure 16: ROC curve for using FFT on the dataset.

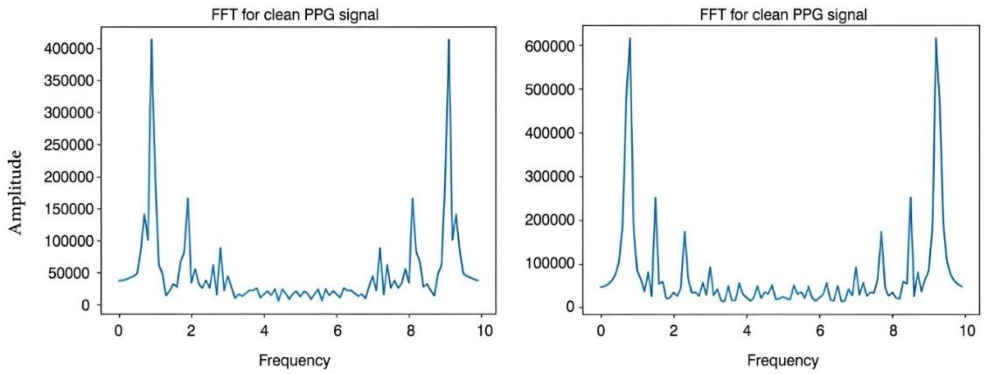

Figure 17: FFT of clean epochs of PPG

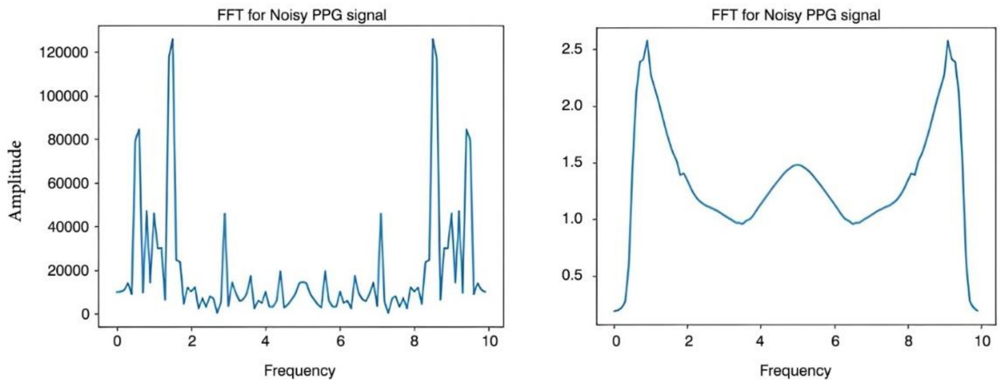

Figure 18: FFT of Noisy epochs of PPG

As can be seen above, the FFT of a clean PPG signal (one that is devoid of noise or artifacts), shows a clear and well-defined peak at the heart rate frequency (usually about 1-2 Hz for a resting individual) and its harmonics. The rest of the frequency spectrum would be rather quiet, indicating that the signal is largely made up of these frequency components (Figure 17). The scenario changes if you take the FFT of a noisyPPG signal (a signal tainted with motion artifacts or other types of noise). While theremay still be a peak at the heart rate frequency, there will also be other peaks corresponding to the noise frequencies (Figure 18). For example, if there is a motion artifact in the signal, you may notice peaks at the frequencies associated with the motion.

5.3 Using Raw Data

Our last approach used the raw data segments of signals for classification and then found the best algorithm for this problem. The algorithm we used for this section is Gradient Boosting Classifier which is an ensemble algorithm.

We gained **78.94%** accuracy and **87.91%** F1 for the classification of raw data without changing the training and testing datasets. The ROC curve is shown in Figure 19.

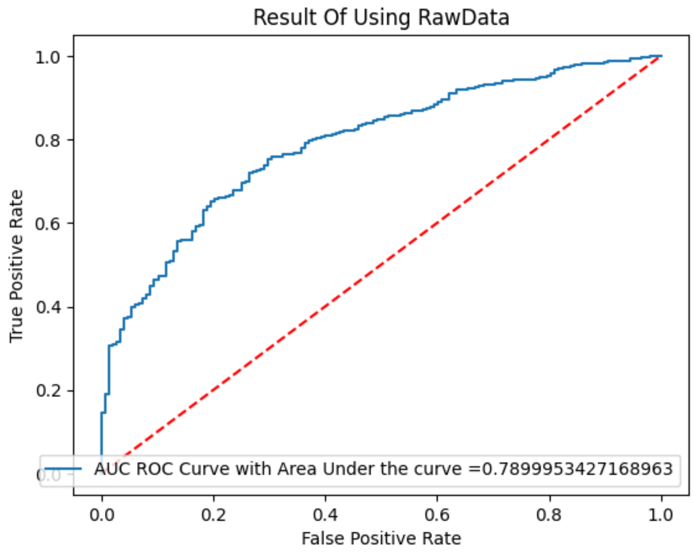

Figure 19: ROC curve of using raw data for classification

The best method with the best results is using feature extraction for all dataset samples.the comparison among them is shown in Table 5-3.

Table 3: Comparison between the methods and their results

Measure	Using Raw Data	Using Feature Extraction	Using FFT
F1	87.91%	93.5%	92.5%
accuracy	78.94%	87.72%	86.4%

5.4 Robustness and generalization

The primary goal of any machine learning endeavor is to discern underlying patternsand structures within that data, not just fit a model to a given set of training data. Theability of a model to generalize, that is, to generate accurate predictions when suppliedwith additional, unseen data, is the ultimate test of its efficacy. The inherent complexity and variety of real-world data pose a hurdle in accomplishing this. The real world frequently presents circumstances that are significantly more complex or diverse than those represented in our training datasets. We used a rigorous technique to test our model's robustness and generalizability. Rather than depending on a singledataset, we used data from five different participants. Because of this variety, we wereable to replicate a more realistic environment in which the model would meet data points it had never seen before during the training phase. The goal was to see if our model could predict outcomes consistently and reliably across these various datasets, showing its robustness and generalizability. Table 5-4 summarizes the results to provide a clear and succinct picture of our model's performance across these datasets.The results in this table provide insight into how well our model adapts to unknown data, which is an important criterion for determining its real-world applicability and reliability.

Table4: Testing the algorithms for unseen datasets.

	Raw Data		FFT of data		Feature extraction of data	
	accuracy	F1	accuracy	F1	accuracy	F1
Subject1	87.11%	92.97%	90%	94.64%	86.88%	92.81%
Subject2	70.66%	82.35%	71.55%	83%	73.55%	84.48%
Subject3	72.88%	82.81%	76.88%	86.52%	76.22%	86.15%
Subject4	86.44%	92.44%	90%	94.55%	89.11%	94%
Subject5	94%	96.87%	96%	97.93	95.12%	97.47%

Chapter 6: Conclusion

Due to the voluminous nature of smartwatch PPG data, the use of expert labeling is out of the question. Therefore, in this project, we used an automated method of labeling noisy segments PPG signal with the help of an accelerometer signal embedded in the smartwatch. The purpose of this project was to detect noisy and clean segments of PPG data which can help us to find the correct health information like BPM and HRV. After acquiring the data from our participants and creating the ground truth labeled dataset, we tested some methods of training and testing data to find out which of them could better detect the noisy and clean epochs. Using statistical features of data and FFT on the dataset gave us promising results for detecting the artifacts in PPG segments. Because the dataset is imbalanced due to being real data from a smartwatch the metric for classification could be F1 instead of accuracy. The F1 of using FFT and statistical features of data for classifications are 92.5% and 93.5% respectively.

6.1 Limitations and further work

We used the accelerometer signal to find noisy epochs. However, using only the accelerometer signal to determine motion and noise artifacts is not sufficient, for many reasons. One of the main problems is that the PPG signal quality can be poor even without arm movement. This can occur due to a myriad of reasons but two primary reasons are bad contact of the PPG sensor to the skin (watch not worn tightly), and darker skin color in some subjects [15]. To detect PPG waveform distortion, time domain features are not that effective because a small change in the PPG time series can lead to a huge deviation in the time domain-based features. As a result, in this study, PPG signal quality has been analyzed in the time-frequency domain. Based on the analysis of the time-frequency spectra (TFS), we want to determine whether the time domain PPG signal is corrupted by MNA or not. For this issue, we can use the variable frequency complex demodulation (VFCDM) and add this method to our projects to detect the noises in segments and label them to have a better dataset. The other issue that we can consider in our project is using an ECG signal [16]. We can use ECG for labeling the dataset beside the Accelerometer signal. We can use R-peak detection in ECG and PPG signals and use RR intervals of both signals for finding artifacts or noise too. Secondly, the accelerometer-based method for finding ground truth eliminates epochs acquired during exercise, which is arguably the most important time to measure HR. It is possible to acquire a clean PPG signal during exercise if the watch is worn correctly. Future work should establish a better way for labelling good/bad epochs/one that does not depend solely on the magnitude of the accelerometer signal.

Bibliography

[1] https://www.mobileindustryreview.com/2016/10/33860.html
[2] https://www.wareable.com/smartwatches/smartwatch-timeline-history-watches
[3] Milad Armani Dehghani, Ki Joon KimKi Joon Kim, (2019). The effects of design, size, and uniqueness of smartwatches: Perspectives from current versus potential users
[4] Denisse Castaneda, Aibhlin Esparza, A review on wearable photoplethysmography sensors and their potential future applications in health care,2019
[5] Jicheng Fu, Shuai Zhang, A Novel Mobile Device-Based Approach to Quantitative Mobility Measurements for Power Wheelchair Users,2021
[6] Fred Shaffer, and J. P. Ginsberg, An Overview of Heart Rate Variability Metrics and Norms, Front Public Health. 2017; 5: 258
[7] Laurie Kelly McCorry, Physiology of the Autonomic Nervous System, Am J Pharm Educ. 2007 Aug 15; 71(4): 78
[8] Kan Wang, Fariba Ahmadizar, Sven Geurts, Heart Rate Variability and Incident Type 2 Diabetes in General Population, the Journal of Clinical Endocrinology & Metabolism, dgad200, April 2023
[9] Florian Custodis MD, Jan-Christian Reil MD, Heart rate: A global target for cardiovascular disease and therapy along the cardiovascular disease continuum, Elsevier2013
[10] Bernhard Grässler, Milos Dordevic, is there a link between heart rate variability and cognitive decline? A cross-sectional study on patients with mild cognitive impairment and cognitively healthy controls, Arq Neuropsiquiatr. 2023 Jan; 81(1): 9–18
[11] K. R. Arunkumar1 · M. Bhaskar, CASINOR: Combination of adaptive filters using single noise reference signal for heart rate estimation from PPG signals, Springer2020
[12] Filipa Esgalhado, Beatriz Fernandes, Valentina Vassilenko, The Application of Deep Learning Algorithms for PPG Signal Processing and Classification, Computers 2021, 10, 158.
[13] Junyung Park, Hyeon Seok, Sang-Su Kim, and Hangsik ShinP photoplethysmograph Analysis and Applications: An Integrative Review, Frontiers2022
[14] Tin Kam Ho, AT&T Bell Laboratories, Random Decision Forests. Proceedings of the 3rd International Conference on Document Analysis

and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282,1995
[15]Syed Khairul Bashar1, Dong Han1, Shirin Hajeb-Mohammadalipour, Atrial Fibrillation Detection from Wrist Photoplethysmography Signals Using Smartwatches,2019
[16] Kaat Vandecasteele, Jesus Lazaro, Evy Cleeren, Artifact detection of wrist photoplethysmograph signals,2018
[17] Shao Hanyu and Chen Xiaohui. 2017. Motion artifact detection and reduction in PPG signals based on statistics analysis. In 2017 29th Chinese control and decision conference (CCDC). IEEE, 3114–3119
[18] Syed Khairul Bashar, Dong Han, Apurv Soni, David D McManus, and Ki H Chon. 2018. Developing a novel noise artifact detection algorithm for smartphone PPG signals: Preliminary results. In 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE, 79–82
[19] Wei-Jheng Lin and Hsi-Pin Ma. 2016. A physiological information extraction method based on wearable PPG sensors with motion artifact removal. In 2016 IEEE international conference on communications (ICC). IEEE, 1–6.
[20] M Raghuram, Kosaraju Sivani, and K Ashoka Reddy. 2016. Use of complex EMD-generated noise reference for adaptive reduction of motion artifacts from PPG signals. In 2016 international conference on electrical, electronics, and optimization techniques (ICEEOT). IEEE, 1816–1820.
[21] Shinsuke Hara, Takunori Shimazaki, Hiroyuki Okuhata, Hajime Nakamura, Takashi Kawabata, Kai Cai, and Tomohito Takubo. 2017. Parameter optimization of motion artifact canceling PPG-based heart rate sensor using cross-validation. In 2017 11th international symposium on medical information and communication technology (ISMICT). IEEE, 73–76
[22] Khawaja Taimoor Tanweer, Syed Rafay Hasan, and Awais Mahmood Kamboh. 2017. Motion artifact reduction from PPG signals during intense exercise using filtered X-LMS. In 2017 IEEE international symposium on circuits and systems (ISCAS). IEEE, 1–4.
[23] Chi-Chin Wu, I-Wei Chen, and Wai-Chi Fang. 2017. An implementation of motion artifacts elimination for PPG signal processing based on recursive least squares adaptive filter. In 2017 IEEE biomedical circuits and systems conference (BioCAS). IEEE, 1–4
[24] Sayeed Shafayet Chowdhury, Rakib Hyder, Real-Time Robust

Heart Rate Estimation from wrist-type PPG Signals Using Multiple Reference Adaptive Noise Cancellation, IEEE2016
[25] ChinmayChakraborty, SubhenduKumarPani, ImplementationofSmart HealthcareSystemsusing AI, IoT, and blockchain, Intelligent Data-Centric Systems
2023, Elsevier
[26] Pierre Geurts · Damien Ernst · Louis Wehenkel, "Extremely randomized trees", Published online: 2 March 2006, Springer Science + Business Media, LLC 2006
[27]J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," Mathematics of Computation, vol. 19, no. 90, pp. 297-301, 1965.
[28] Eugene N. Bruce, Biomedical Signal Processing and Signal Modeling, November 2000
[29]YiruWang, WeihaoGan, JieYang, DynamicCurriculumLearningforImbalancedDataClassification, arXiv:1901.06783v2 [cs.CV] 15 Aug 2019
[30]NiteshV.Chawla, KevinW.Bowyer, Lawrence.Hall, 'SMOTE: SyntheticMinorityOver-samplingTechnique' , Journal of Artificial Intelligence Research16(2002)321–357
[31]https://www.geeksforgeeks.org/bagging-vs-boosting-in-machinelearning/
[32]https://www.geeksforgeeks.org/bagging-vs-boosting-inmachine-learning/
[33]https://www.geeksforgeeks.org/bagging-vs-boosting-inmachine-learning/
[34]https://www.turing.com/kb/random-forest-algorithm