SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage: https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

ECG Biometric Verification Incorporating Different Physiological Conditions

Siti Nurfarah Ain Mohd Azam[1], Khairul Azami Sidek[1,*], Alaa Kamal Yousif Dafhalla[2]

[1] Department of Electrical and Computer Engineering, Faculty of Engineering, International Islamic University Malaysia, Selangor, Malaysia
[2] Department of Computer Engineering, College of Computer Science and Engineering, University of Ha'il, Kingdom of Saudi Arabia

ARTICLE INFO

Article history:
Received 22 November 2023
Received in revised form 1 April 2024
Accepted 21 August 2024
Available online 18 September 2024

Keywords:
Biometric; ECG; MODWT

ABSTRACT

The liveness detection criteria of biological signals have become one of the reasons it has been introduced as an ideal biometric recognition system. Electrocardiogram (ECG) is one of the biological signals that records the rhythms of human's heart in the form of PQRST waves proves the uniqueness of the ECG itself making it suitable to be applied as biometric mechanisms. Previous research had shown the success of proving ECG as a biometric modality however most experimentations were done in normal conditions. Thus, to improve the current research, this work proposed a robust biometric identification by introducing ECG signals incorporating various physiological conditions. After the data collection of cycling, walking, climbing stairs, and jogging, the signals are pre-processed by using MODWT to remove unwanted noises produced during data collection process. Then, Pan Tompkins algorithm is used to segment the QRS complexes. The segmented signals are overlapped and align with each other to observe its pattern. Next, the QRS waveform is classified by using various class of SVM by considering two factors which are same physiological conditions and different physiological conditions. The subjects are compared between same and different physiological condition to validate the proposed method. The results show that the precision achieved up to 100%. In average, Gaussian SVM gives highest precision when compared to other type of SVM classifiers suggesting that Gaussian SVM is the most appropriate to be applied for person identification. Thus, the proposed method proves that biometric recognition can be performed regardless of different physiological conditions and can be applied in real life scenarios.

1. Introduction

1.1 Background Study

Biometric is described as "the automated recognition of individual based on their behavioural and biological characteristics" [1]. In other words, it is a method of person authentication by computing and approximating one physiological trait. There are many types of existing biometric modalities that have been actively used nowadays such as fingerprint, signature, and voice. These

* *Corresponding author.*
E-mail address: azami@iium.edu.my

https://doi.org/10.37934/araset.51.2.97110

types of biometrics are categorized as the first generation of biometrics where it mostly used physical body parts. However, there are several limitations from the first-generation biometrics. For example, fingerprints can be duplicated which could cause fraud. Voice is sensitive to noise that might reduce the identification accuracy and signature can be copied without consent. Due to these constraints, a second generation of biometrics is proposed by applying biological signals in order to get a grip on these drawbacks. Biological signal is a promising biometric feature because of its ability to provide high accuracy and give more security as it has the liveness detection criteria. Therefore, it is hard to replicate, and the data is secured. Electroencephalogram (EEG), Electrocardiogram (ECG) and Photoplethysmogram (PPG) are a few examples of biological signals. For this study, ECG signal will be applied as a biometric mechanism.

ECG is the measure of the heart's activity where it picks up the electrical signal that is formed by the polarization and depolarization of cardiac tissues and translates them into PQRST waveform as shown in Figure 1. The waveform provides essential information of the heart conditions where ECG is commonly used in the medical area to identify heart diseases. The principle of ECG is where it distinguishes and amplifies the tiny electrical signal that fluctuates on the skin which is caused when the heart muscle depolarizes during each heartbeat.

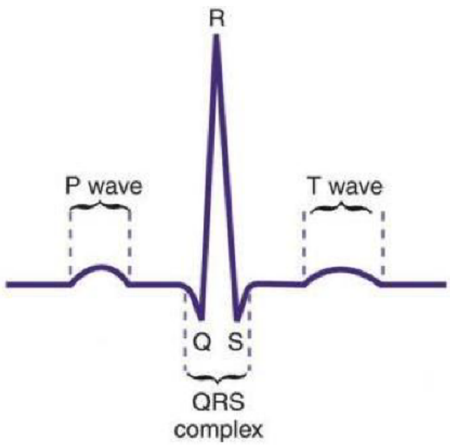


Fig. 1. The morphology of ECG signals

Earlier research has shown the validity of applying ECG signals as security mechanism on different individuals. However, this work has identified the one of the research problems that is to increase the user acceptability regarding the capability of ECG to be applied as a biometric feature. Another concern is that previous investigations have been focusing on person identification technique in normal condition without highlighting the fact of different physiological conditions. In real-life scenarios, people are making movement and consistently doing activities such as walking down the stairs, running and jumping. Thus, in order to increase trust in people, this work will focus on the feasibility of applying ECG signals as a biometric feature by considering different physiological conditions. The outcome of this research will enable person identification using ECG signal under different physiological conditions.

1.2 Related Works

To perform biometric, Aziz *et al.,* [2] proposed an Empirical Mode Decomposition (EMD) to remove noise on the raw ECG signal as a pre-processing technique. A total of 14 subjects consists of 8 males and 6 females participate in the study where the data is obtained by using BIOPAC. All the subjects are at rest and the data are sampled at the sampling frequency of 1000 Hz. The acquired data is pre-processed using EMD where it decomposed the raw ECG signal into multi sub-components known as Intrinsic Mode Functions (IMF). The simulation shows that only IMF1 and IMF2 consist of real information of the ECG signals whereas the other are noises. Then, IMF1 and IMF2 are combined to make a new signal. The new signal then undergoes feature extraction process where the signals are distinguished by applying Shannon Energy, skewness, variance, occupied bandwidth and median frequency. These selected features are then tested with different classification approaches. The results show that Cubic SVM managed to achieved the highest classification accuracy of 98.72% suggesting that the proposed method is suitable to perform biometric authentication using the ECG signals.

Alotaiby *et al.,* [3] proposed a human identification system based on the common spatial pattern (CSP) feature extraction technique by using ECG. A total of 200 subjects were collected from Physikalisch-Technische Bundesanstalt (PTB) database which consists of 52 healthy subjects and 148 non-healthy subjects. CSP works by differentiating subject-related ECG signal with non-subject-related ECG signal. For pre-processing, Hilbert transform is used where it constructs an additional channel to fulfill the requirement of multichannel information for CSP algorithms. After pre-processing, the signals are segmented into six partitions that are 1,3,5,7,10 and 15 seconds. The segmented signals are then classified by using SVM with a Radial Basis Function (RBF). The Identification Rate (IR) and Equal Error Rate (EER) are used to evaluate the performance of the proposed technique. The outcome shows that the proposed system managed to achieve high IR of 95.15% and EER of 0.1 on single limb-based lead (I) and IR of 98.92% and EER of 0.08 using single chest-based lead (V3). However, both of the study conducted by Aziz *et al.,* [2] and Alotaiby *et al.,* [3] can be further improved by considering subjects with different physiological conditions so that the system can be fully utilized in various fields such as banking.

Hadiyoso *et al.,* [4] proposed using Hjorth Descriptor (HD) and Sample Entropy (SE) as feature extraction method for the ECG biometric identification system. A total of 10 subjects are involved in the study where the readings are taken during normal conditions. The experiment is conducted by using one-lead ECG device attached on the arm approximately for 60 seconds with a sampling frequency of 100 Hz. The feature extraction coefficient is obtained by using HD and SE which the signal complexity parameters based on the ECG is developed. The validity of proposed technique is classified by using Linear SVM, Cubic SVM, Quadratic SVM and Gaussian SVM and 10-fold cross-validation is applied. The results show that the highest accuracy obtained is 93.8% by using HD with Gaussian SVM. Compared to SE technique, the highest classification accuracy achieved is 93.8% (Gaussian SVM) which is low. Thus, HD is a better feature extraction method compared to SE approach. However, the accuracy value can be increased if the filtering process is added to the pre-processing step as noise produced by the ECG is one of the factors that has reduced the accuracy.

Bastos *et al.,* [5] proposed double authentication process by using PPG and ECG signals on wearable devices. There are three stages involved which are template storage, user test and user correlation. A total of 53 subjects were involved in the study where the data are taken from BIDMC database. For pre-processing step, Chebyshev I low pass filter is used for PPG and Maximum Overlap Discrete Wavelet Transform (MODWT) IV filter is applied for ECG signal. After filtering, the highest peaks are identified and based on the reference peaks the signals are segmented into several windows. Based on the windows, the signals are overlapped and align with each other. Then, Cross-Correlation Function (CCF) is used to create storage template where the average correlation among the overlapped signals is calculated. Based from the CCF, the authentication threshold is measured to be used as reference model to identify the user. The results show that, PPG gives the accuracy of 99.98% meanwhile ECG gives the accuracy of 88.79%.

The remaining sections of this paper are separated into three sections. Section 2 clarifies on the technique of study that explains the process of data collection, pre-processing, feature extraction of QRS and signal classification. Then, in Section 3, results of the experimentation are discussed. Lastly, based on the results, conclusion is elaborated in Section 4.

2. Methodology

This section elaborates the method used in the study which is divided into four sections as shown in Figure 2 that are data acquisition, pre-processing stage, feature extraction and signal classification.

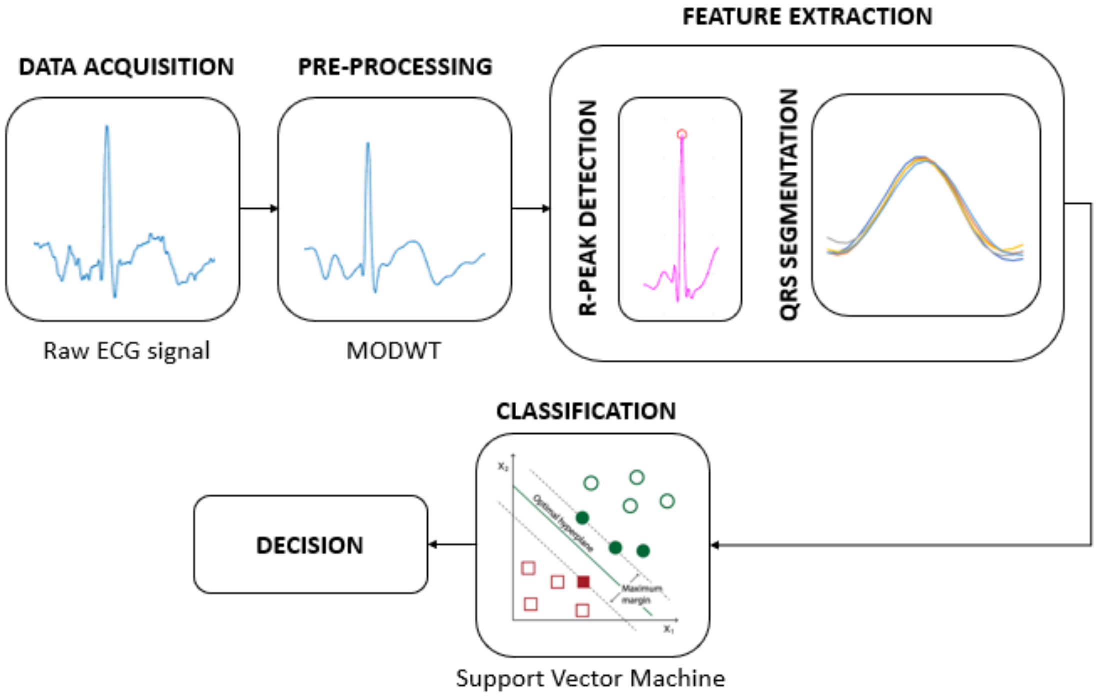


Fig. 2. The proposed identification system

2.1 Signal Acquisition

In the study, a total of 10 subjects were involved where the ECG data are collected from Mobile Health (mHealth) [6] dataset performing various physical activities. To collect ECG, the sensor is positioned on the chest as shown in Figure 3 to evaluate the effect of conducting different exercises on the ECG signal.

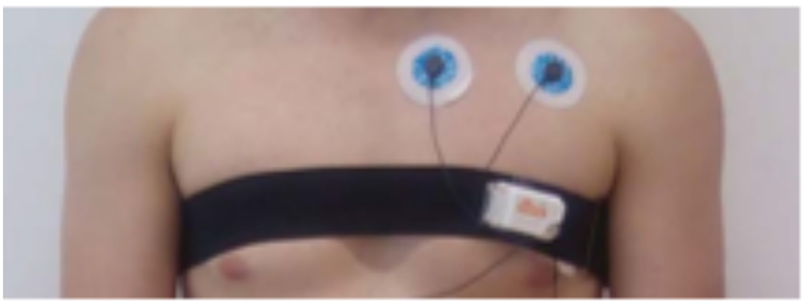


Fig. 3. Shimmer2 (BUR10) with 2 lead sensors for ECG recording [7]

In real life scenario, human body is consistently making movements. Body movements will cause instability to ECG waveform. Therefore, to adapt daily life scenarios, four different physical activities are considered in this work which are cycling, walking, climbing stairs and jogging. Each activity was collected in an out-of-lab environment where there are no restrictions on how the exercises are conducted. The data were recorded at a sampling rate of 50 Hz for duration of 60 seconds.

2.2 Pre-Processing

Raw ECG signal commonly contaminated with noise due to the interference from uncontrolled movement while taking the ECG recording. The presence of noise will give an undesirable output thus reducing the accuracy of the proposed technique. Hence, it is very important to have a noise-free signal in order to achieve an optimal result. In the study, Maximal Overlap Discrete Wavelet Transform (MODWT) is used. The MODWT is an undecimated wavelet transform where the concept is nearly similar to the Discrete Wavelet Transform (DWT) with few advantages. MODWT is able to convert signals into coefficients of details and approximation unaffected by any shift efficiently. In addition, MODWT has the advantage of it able to define at any signal length which means there are no restrictions unlike DWT and is not sensitive on the choice of the origin of the signal [8]. Besides that, it also has the ability of handling arbitrary sample sizes [9].

Next step is to select an appropriate wavelet which has the similar signal feature as ECG waveform. By referring to Daubechies family of wavelet, db4 wavelet provided the good signal output because the wavelet has the almost identical shape as the ECG morphology as shown in Figure 4. The signal decomposed down to level 4. MODWT filters the noise coefficients of the input signal resulting in smoother and cleaner ECG waveform. This process will separate the ECG waveform from unwanted disturbance ensuring that other interfering signals are absent, and error can be minimized.

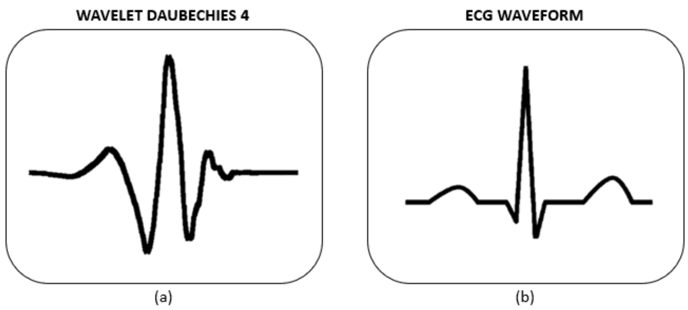


Fig. 4. (a) Wavelet Daubechies 4 [10] (b) ECG morphology

2.3 Feature Extraction

One complete ECG signal consists of 5 peaks which are P, Q, R, S and T. However, for feature extraction the QRS segment was chosen in the analysis as shown in Figure 5 as it is most less affected by cardiac irregularities, noise and artifacts as shown in previous works. The algorithm studied by Pan and Tompkins [11] is applied to segment the QRS waveform and R peaks will be used as the reference point. As we can observe from Figure 5, the R wave corresponds to the highest peak in the ECG signals as a result from the activity of the ventricular. Segmentation is important in order to identify the location of Q, R and S peaks. The R peaks are true if the value of the peaks is higher than the threshold. Then, the value of Q and S peaks are obtained by creating a window in reference to the R peaks.

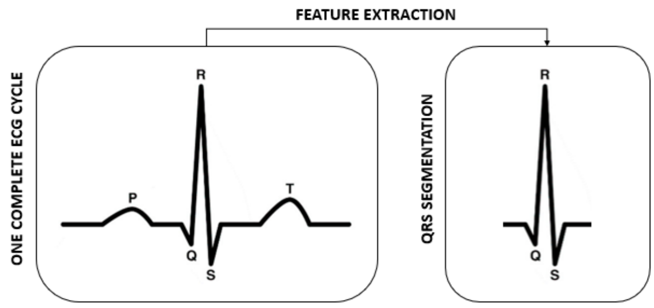


Fig. 5. Segmentation of QRS waveform

2.4 Classification

Support Vector Machine (SVM) is a supervised machine learning model which is applicable for regression, classification and outlier detection on linear and non-linear data [12,13]. An SVM constructs a hyperplane or set of hyperplanes which also known as the decision boundaries that help to categorize the data points by constructing a straight line between two classes [14] as depicted in Figure 6.

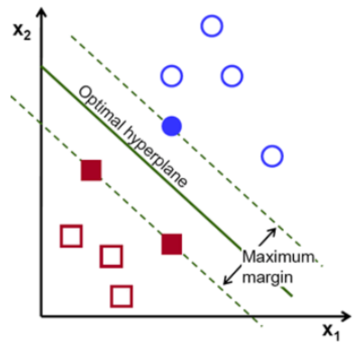


Fig. 6. Support vector machine [15]

Optimal hyperplane is obtained by taking the support of two other hyperplanes that are parallel and equidistant from both sides [12,16]. The support hyperplanes are also called support vectors. Maximum margin is important in order to reduce the chance of misclassification. Hyperplane with higher margin value gives more assurance in data accuracy [15]. SVM gives more accurate outcomes as their ability to handle small and complex datasets [17]. Besides that, SVM also results in better performance with less computational power and significant accuracy compared to other types of machine learning algorithms [14]. The overall architecture of the support vector machine is shown in Figure 7.

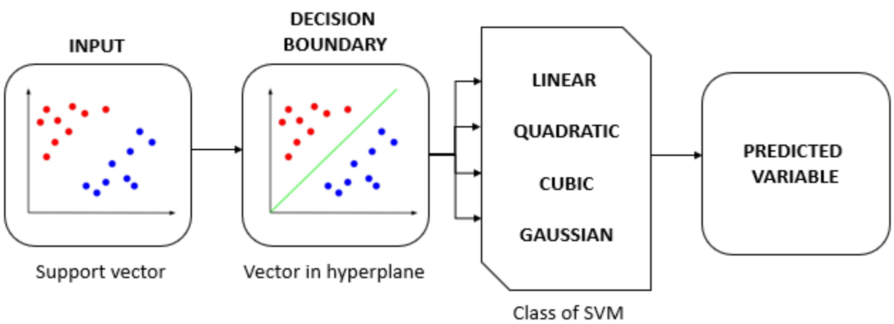


Fig. 7. General architecture of support vector machine [15]

In the study, the proposed technique is evaluated by two classification models which are accuracy and training-testing. Accuracy is a metric that measures how well a classification model correctly predicts the class labels of the test instances. It is calculated as the ratio of the number of correct predictions to the total number of predictions made by the model where it provides an overall assessment of the model's performance in terms of correct predictions. Whereas for training-testing, the dataset is split into two subsets that are the training set and the testing set. The training set is used to train the model and the testing set is used to validate the model's performance by measuring the precision.

3. Results

During ECG data collection process, there will be unpredictable movement as the subjects are conducting simple exercises. These movement apparently will result in the presence of noise and unwanted disturbance on the ECG signals recording. It is well known that ECG signals are sensitive to disturbances and due to the noise existence, the ECG signals will show patterns that are not intrinsic and resulting in less accurate data for analysis. Thus, these non-intrinsic patterns or also known as trends must be removed in order to increase the accuracy of the ECG data to perform biometric. As can be observed in Figure 8(a), the original ECG signals show a baseline shift. By computing the MATLAB from MathWorks [18] command detrend, the baseline shift is corrected as shown in Figure 8(b). By detrending the signal, the trends are removed, and the accuracy of data analysis will be increased as the baseline shift is properly shifted.

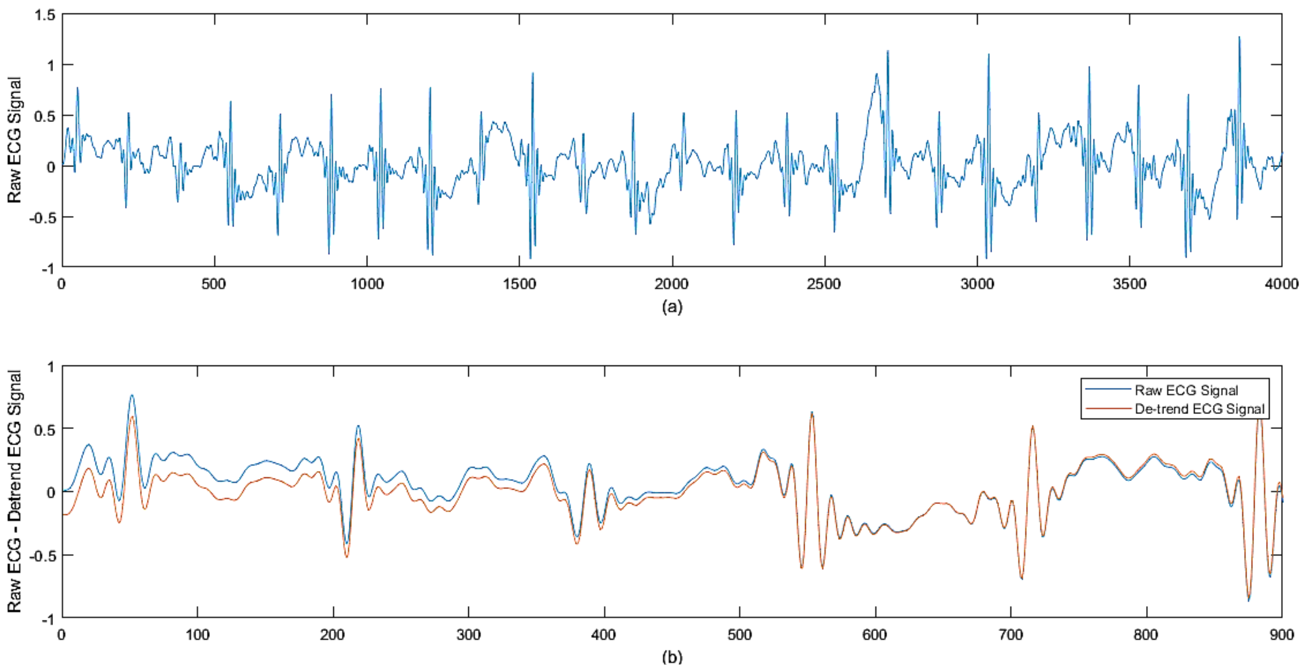


Fig. 8. (a) Raw ECG signal and (b) Zoom-in detrend ECG signal for Subject 1 (cycling)

As mentioned before, the ECG signals are sensitive to movement, and it is very common that raw ECG signals are contaminated with noises. It can be clearly seen in Figure 8(a) that the PQRST waveform of the raw ECG signals is polluted with noises. The presence of noise will reduce the accuracy of the proposed technique. Therefore, pre-processing process is important to reduce as much signal fluctuations as possible in order to enhance the accuracy of the proposed method.

Noise from the raw ECG signal is filtered out by using the db4 wavelet with Maximal Overlap Discrete Wavelet Transform (MODWT) denoising and a level 4 decomposition is applied to it. In the study, the soft thresholding is used. Figure 9(a) shows the combination of two signal which are raw ECG and filtered ECG signal from Subject 3 (cycling). Based on the figure, the filtered signal has become smoother compared to the noisy signal. Figure 9(b) illustrates a better diagrams of ECG signal that has been filtered using MODWT. As can be seen in the figure, the shape of ECG signal is preserved, and the unwanted noise has been successfully filtered out. Thus, MODWT is a good filtering technique that can be applied on noisy ECG signal.

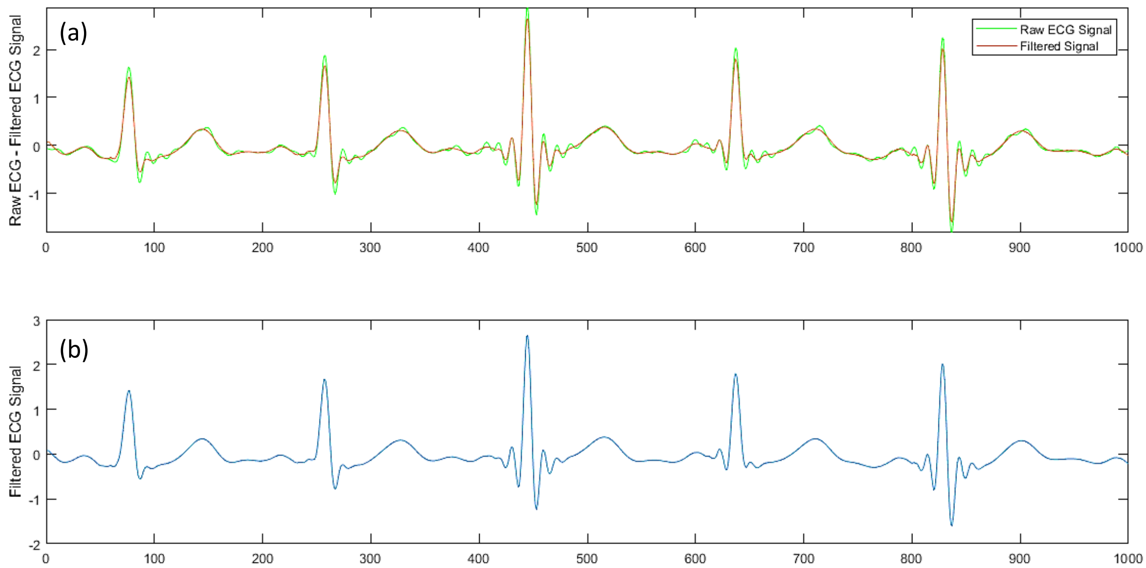


Fig. 9. (a) Zoom in raw ECG - filtered ECG signal of Subject 3 (cycling) and (b) Filtered ECG signal

Next, is the feature extraction stage. The purpose of QRS segmentation is to compare the pattern and observed its characteristics based on time and morphology. Based on the previous research conducted by D'Aloia *et al.,* [19], QRS segment is less affected by noise and muscle artefacts thus it will help in enhancing the accuracy of data analysis. The segmentation process started by identifying the R peak (marked by 'o') of the ECG signal as the reference point as shown in Figure 10. Based on the reference point, window is created where a total of 21 points were selected to the left and right of the waveform.

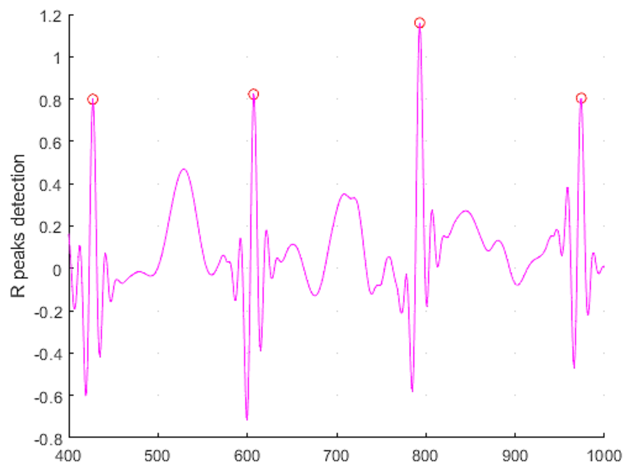
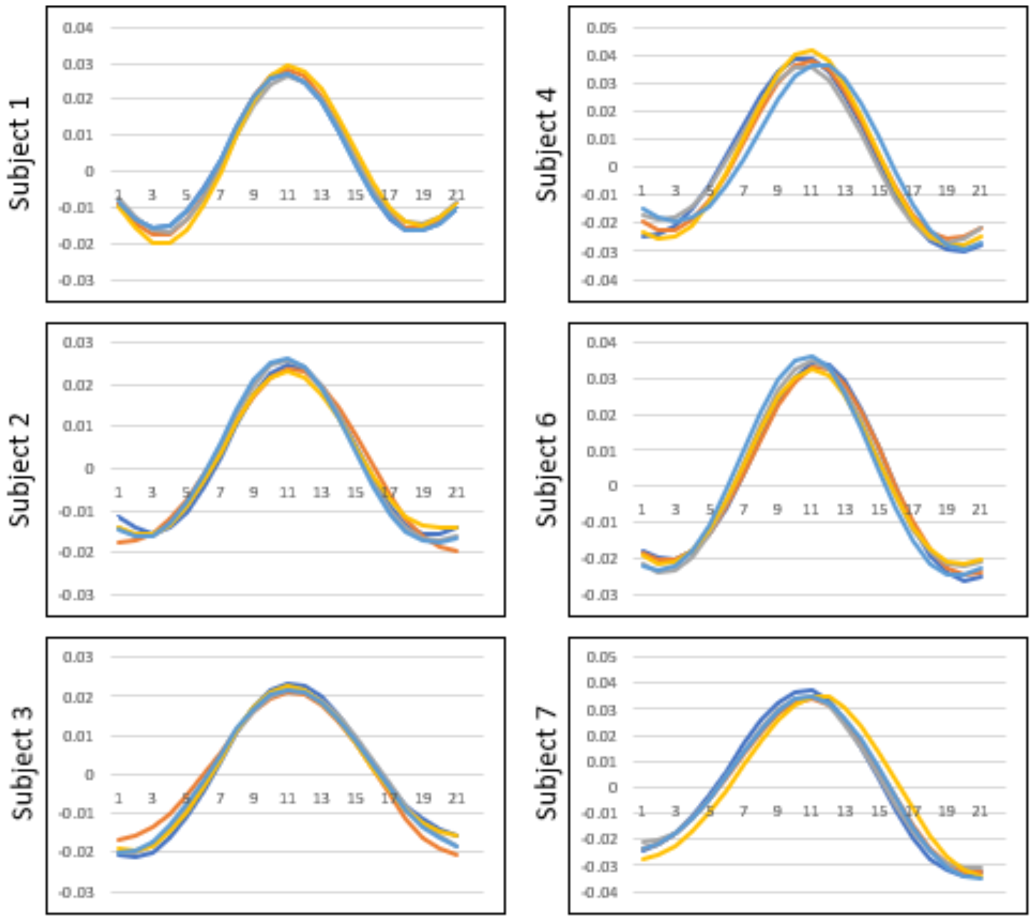


Fig. 10. Zoom in R peaks detection for Subject 1 (walking)

Then, the signals were overlapped and aligned with each other to observe the pattern as shown in Figure 11. Each ECG samples consists of 10 QRS complexes. Figure 11 demonstrate the result of QRS segmentation for Subject 1, Subject 2 and Subject 3 during cycling activity and Subject 4, Subject 6, and Subject 7 during walking exercise. One of the criteria of biometric is distinctiveness where each person must have their own unique characteristics [20]. It can be seen that each subject has their own unique pattern of QRS shape thus it can presume that biometric by using ECG signals considering different physiological conditions is achievable.



(a) (b)

Fig. 11. Feature extraction of QRS segmentation for (a) Cycling and (b) Walking subjects

Mean value for each subject is calculated based on the QRS samples. The purpose is to observe information received from each dataset. The mean is the sum of the numbers in a data set divided by the total number of values in the data set [21]. In addition, the overall idea of the data can be obtained from mean value. Table 1 demonstrate the results of the calculation.

The result of experimentation shows that each subject has different value of mean. Figure 12 show better illustration of Table 1 for better understanding. As can be seen in Figure 12, every subject has a unique identifier that separate them from other subject. This satisfies one of the characteristics of biometric that is inimitable. The experiment is continued by performing classification towards the subjects. The investigation is separated into two categories. The first part of experiment is each subject performance are evaluated using the same physiological condition. For the second part of the experimentation, the subjects are classified with different physiological conditions. The details of the experiment and results are elaborated in the next sub-section.

Table 1
Mean value of QRS segmentation during cycling condition

Subject	Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8	Sub 9	Sub10
Point 1	-0.0088	-0.0144	-0.0195	-0.0169	-0.0050	-0.0159	-0.0222	-0.0049	-0.0033	-0.0151
Point 2	-0.0139	-0.0158	-0.0193	-0.0165	-0.0066	-0.0178	-0.0216	-0.0063	-0.0035	-0.0190
Point 3	-0.0171	-0.0157	-0.0173	-0.0146	-0.0070	-0.0181	-0.0188	-0.0068	-0.0032	-0.0212
Point 4	-0.0168	-0.0133	-0.0136	-0.0107	-0.0051	-0.0160	-0.0139	-0.0059	-0.0022	-0.0203
Point 5	-0.0131	-0.0088	-0.0084	-0.0052	-0.0014	-0.0116	-0.0072	-0.0037	-0.0006	-0.0156
Point 6	-0.0066	-0.0026	-0.0021	0.0015	0.0036	-0.0052	0.0010	-0.0004	0.0017	-0.0075
Point 7	0.0019	0.0045	0.0046	0.0088	0.0094	0.0024	0.0097	0.0036	0.0043	0.0033
Point 8	0.0112	0.0119	0.0110	0.0159	0.0155	0.0104	0.0179	0.0078	0.0069	0.0155
Point 9	0.0197	0.0185	0.0165	0.0219	0.0209	0.0176	0.0248	0.0117	0.0092	0.0274
Point 10	0.0257	0.0231	0.0204	0.0260	0.0240	0.0229	0.0294	0.0143	0.0109	0.0370
Point 11	0.0278	0.0247	0.0221	0.0271	0.0240	0.0253	0.0308	0.0151	0.0117	0.0421
Point 12	0.0257	0.0231	0.0214	0.0252	0.0209	0.0246	0.0288	0.0139	0.0113	0.0418
Point 13	0.0202	0.0190	0.0187	0.0206	0.0152	0.0211	0.0239	0.0109	0.0096	0.0360
Point 14	0.0123	0.0129	0.0142	0.0136	0.0076	0.0154	0.0164	0.0066	0.0068	0.0255
Point 15	0.0031	0.0054	0.0085	0.0053	-0.0010	0.0082	0.0074	0.0015	0.0030	0.0120
Point 16	-0.0055	-0.0022	0.0023	-0.0034	-0.0094	0.0006	-0.0020	-0.0036	-0.0013	-0.0027
Point 17	-0.0119	-0.0087	-0.0038	-0.0114	-0.0162	-0.0062	-0.0109	-0.0079	-0.0057	-0.0161
Point 18	-0.0153	-0.0133	-0.0090	-0.0177	-0.0207	-0.0112	-0.0182	-0.0107	-0.0098	-0.0263
Point 19	-0.0156	-0.0157	-0.0131	-0.0218	-0.0228	-0.0143	-0.0232	-0.0120	-0.0132	-0.0321
Point 20	-0.0135	-0.0166	-0.0159	-0.0237	-0.0233	-0.0158	-0.0258	-0.0120	-0.0156	-0.0335
Point 21	-0.0096	-0.0161	-0.0177	-0.0238	-0.0226	-0.0160	-0.0262	-0.0111	-0.0171	-0.0312

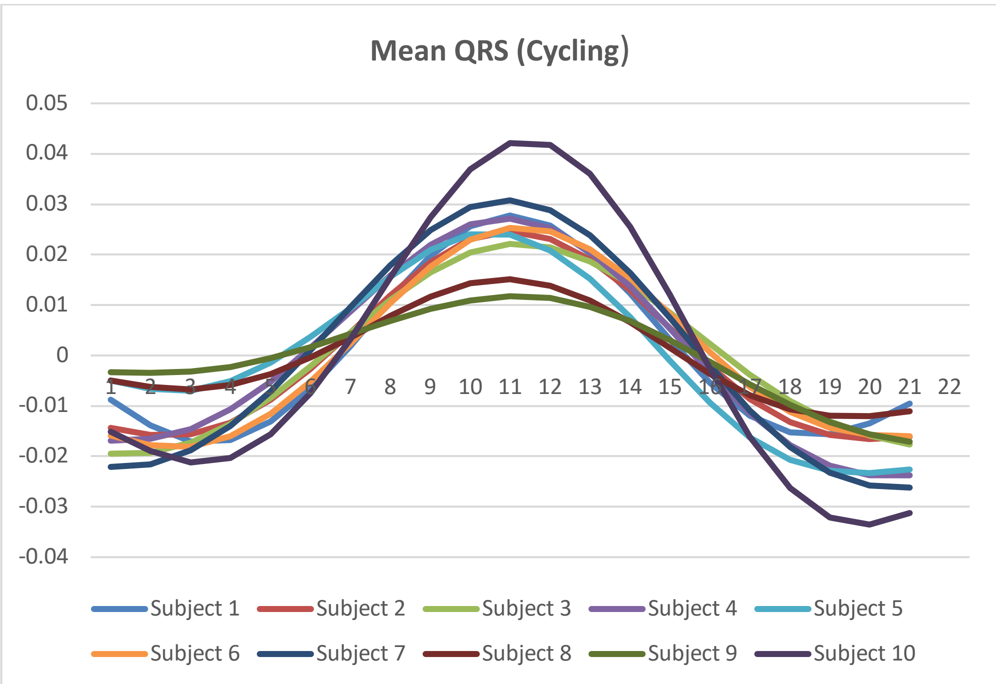


Fig. 12. Mean QRS segmentation of subject during cycling

3.1 Classification Accuracy of the Same Physiological Condition for Each Subject

In this sub-section, the classification accuracy of the same physiological condition for each subject is discussed. There are four conditions and comparisons will be performed which are cycling against cycling (C-C), walking against walking (W-W), climbing stairs against climbing stairs (CS-CS) and jogging against jogging (J-J). Table 2 shows the classification results of the proposed system by using multi-class of SVM which are Linear SVM (L-SVM), Quadratic SVM (Q-SVM), Cubic SVM (C-SVM) and Gaussian SVM (G-SVM). The purpose of classifying the signal by using multi-class SVM is to evaluate the best classifier for the proposed technique. In this work, 10-fold cross-validation is applied. As mentioned in the previous section, the proposed technique is evaluated by two classification model which are accuracy and training-testing.

Table 2
Classification accuracy of QRS segmentation of same physiological condition between subject

Classification	Accuracy (%)				Training -Testing (%)			
	L-SVM	Q-SVM	C-SVM	G-SVM	L-SVM	Q-SVM	C-SVM	G-SVM
C - C	80	93	86	93	100	100	100	100
W - W	82.5	85	80	92.5	100	100	100	100
CS - CS	86.7	94.2	86.7	80	90	91	93.2	100
J - J	68.1	89.3	71.9	85.6	86.7	83.3	83.3	96.7
Average	79.33	90.34	81.15	87.78	94.18	93.56	94.13	99.18

*Note: C = cycling, W = walking, CS = climbing stairs, and J = jogging

Comparing the same physiological condition, it is expected that the classification precision is high as the morphology of ECG signal will be similar. Based on the result of the experimentation, for accuracy measurement, the highest percentage of precision is 94.2% (Quadratic SVM) for CS-CS condition and J-J condition gives the lowest precision of 68.1% (Linear SVM). For average, Quadratic SVM (90.34%) gives the highest precision of accuracy when compared to Gaussian SVM (87.78%), Cubic SVM (81.15%) and Linear SVM (79.33%). This result recommended that Quadratic SVM is a good classifier to measure accuracy for person identification in same physiological condition. Quadratic SVM show high accuracy as it can effectively handle datasets that are not linearly separable, allowing for more flexibility in capturing complex relationships between features and labels [22]. According to the result, Linear SVM and Cubic SVM will be the least favourable classifier as the precision obtained are low for all conditions.

Next, for training-testing, the precision achieved is up to 100% for all classifiers during C-C and W-W conditions. The lowest is 83.3% by using Quadratic SVM and Cubic SVM during J-J condition. High precision of accuracy is due to the consistency of the heart rate while conducting exercises and low precision is the result of inconsistency of the heart rate during activity. In addition, the outcome shows the precision more than 90% of precision in average indicating the preciseness of ECG signal in person recognition. Gaussian SVM demonstrate highest precision of 99.18% when compared to Linear SVM (94.18%), Quadratic SVM (93.56%) and Cubic SVM (94.13%). The results suggested that in order to achieve high precision in person verification, Gaussian SVM is a suitable classifier. High precision in Gaussian SVM because it has the regulation property that finds the best balance between maximizing the margin and minimizing the training error [23]. Therefore, high percentage of precision shows the capability of the proposed system for person identification regardless of physiological conditions.

Among all of the physiological conditions, cycling and walking give the higher precision compared to climbing stairs and jogging activities. The result of this condition is because the probability on the consistency of the physical movement while conducting the experimentation. Consistency of the movement gives the stability of the heart rate thus developing constant shape of ECG signals. Jogging shows the least precision rate. This is because jogging requires more physical movement while performing the exercise. The stamina of the subject also influenced the result of the experimentation.

Subjects with good stamina will give constant heart rate where participants with lower stamina show less stable heart rate.

3.2 Classification Accuracy of the Different Physiological Condition for Between Subject

This sub-section discussed on the classification accuracy of different physiological condition between subjects. Six comparisons will be evaluated that are cycling against walking (C-W), cycling against climbing stairs (C-CS), cycling against jogging (C-J), walking against climbing stairs (W-CS), walking against jogging (W-J) and climbing stairs against jogging (CS-J). The motivation is to demonstrate that the ECG signal of a person is permanence regardless of any physiological state thus proving ECG is robust as a biometric feature. Table 3 shows the classification results of the proposed technique.

Table 3
Classification accuracy of QRS segmentation of different physiological condition between subject

Classification	Accuracy (%)				Training -Testing (%)			
	L-SVM	Q-SVM	C-SVM	G-SVM	L-SVM	Q-SVM	C-SVM	G-SVM
C - W	71.8	85.9	83.5	95.7	88.9	93.9	88.9	97.9
C - CS	71.7	90	77.8	98	90	100	90	100
C - J	68.1	69.4	70.8	96.4	75	89.2	82.5	98.5
W - CS	81.2	82.4	94.7	90.8	86.7	87.8	87.8	96.7
W - J	67.2	68.7	68.7	88.7	82.9	97.1	82.9	94.6
CS - J	65.3	72.2	94.7	89.8	80	80	92.5	91.1
Average	70.88	78.1	81.7	93.23	83.91	91.33	87.43	96.47

*Note: C = cycling, W = walking, CS = climbing stairs, and J = jogging

According to the outcome, C-CS combination samples achieved the highest precision of 100% in training-testing method by using Quadratic SVM and Cubic SVM compared to other groupings. This result might be due to the similarity of body movement level and the heart pace of the subjects during activity thus producing almost identical ECG morphology. Overall performance shows high precision of 90% and above which already satisfies the biometric requirement except for W-J combination (88.7%) by using Gaussian SVM in accuracy method. Decrease in classification performance may cause by significant body movement level. Walking is a slower pace activity when compared to jogging thus resulting in slightly different ECG waveform and heart rate. Activity with the nearly identical physical movements such as cycling and jogging will give high precision while different pace of body motion activity will result in lower precision of classification.

As can be observe in the table, different combination of activity will give different results in maximum precision. For example, Gaussian SVM gives the highest precision in accuracy for C-CW (95.7%), C-CS (98%), C-J (96.45) and W-J (88.7%) subjects. Next in line is Cubic SVM. Cubic SVM show the highest precision in accuracy for W-CS (94.7%) and CS-J (94.7%). However, in training-testing result, Quadratic SVM is the second best in precision after Gaussian SVM. In accordance with the outcome, Gaussian SVM is suitable to be used as classification technique. Based on the average, Gaussian SVM shows the highest precision for accuracy with percentage of 93.23% and training-testing method gives the precision value of 96.47% for different physiological condition between subjects. High precision result in Gaussian SVM is because the Gaussian kernel has the ability to minimizes both the estimation and approximation errors of a classifier thus increase the classification accuracy [24]. In case of absence of Gaussian SVM, Cubic SVM is the next choice for accuracy and Quadratic SVM for training-testing.

Based on the results in Table 2 and Table 3, it is suggested that Gaussian SVM is the best classifier to be applied for person identification. Although Quadratic SVM also shows high precision in accuracy, however it is important to standardised using the same classifier. The reason is by using the same classifier will ensure that the accuracy of the data is consistent. In accordance of the results, it can be generalized that, biometric by using ECG signal can be performed on different physiological conditions.

4. Conclusion

In this work, an ECG biometric verification method considering different physiological condition is investigated. A total of 10 subjects were involved in the study where the subjects were conducting four different activities which are cycling, walking, climbing stairs, and jogging. The raw ECG data is filtered using MODWT to remove noise. After that, the clean signal undergoes feature extraction process by applying Pan Tompkins algorithm. The R-peak is detected as a reference point. Based on the reference point, the QRS waveform is segmented. Then, the QRS segments are overlapped with each other to observe its pattern. The study discovered that each subject has different shape of QRS signal that distinguish one person to another which make it as a unique identifier. Later, multi-class of SVM is used to perform classification. The experiment is divided into two parts where the first part involved comparison between the same physiological condition and the second section intricate comparison between different physiological condition for each subject.

For the first part of the experimentation, it is expected to obtain high precision in classification. In average, Quadratic SVM gives the precision of 90.34% for accuracy and Gaussian SVM show 99.18% precision in training-testing. Then, comparison between different physiological condition results demonstrate a slightly lower precision compared to same physiological condition. This might be due to body movement level. For accuracy and training-testing, Gaussian SVM gives highest precision of 93.23% and 96.47% when compared to other SVMs. Therefore, according to the outcome, Gaussian SVM is the best classifier to implement person verification regardless of physiological condition because of the consistently of performing high classification precision. The contribution of the study is the ECG biometric can be performed not only on subjects on static condition but also on subjects that are moving. In addition, the proposed method is less complexity however managed to obtain high identification accuracy. For future work, the study can be improved by increasing the number of samples. The performance of the proposed method should be maintained regardless of any physiological conditions.

References

[1] International Organization for Standardization. "Information technology — Biometrics — Overview and application." *ISO/IEC TR 24741* (2018).

[2] Aziz, Sumair, Muhammad Umar Khan, Zainoor Ahmad Choudhry, Afeefa Aymin, and Adil Usman. "ECG-based biometric authentication using empirical mode decomposition and support vector machines." In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 0906-0912. IEEE, 2019. https://doi.org/10.1109/IEMCON.2019.8936174

[3] Alotaiby, Turky N., Saleh A. Alshebeili, Latifah M. Aljafar, and Waleed M. Alsabhan. "ECG-based subject identification using common spatial pattern and SVM." *Journal of Sensors* 2019 (2019). https://doi.org/10.1155/2019/8934905

[4] Hadiyoso, Sugondo, Suci Aulia, and Achmad Rizal. "One-lead electrocardiogram for biometric authentication using time series analysis and Support Vector Machine." *International Journal of Advanced Computer Science and Applications* 10, no. 2 (2019). https://doi.org/10.14569/IJACSA.2019.0100237

[5] Bastos, Lucas, Thais Tavares, Denis Rosário, Eduardo Cerqueira, Aldri Santos, and Michele Nogueira. "Double authentication model based on ppg and ecg signals." In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 601-606. IEEE, 2020. https://doi.org/10.1109/IWCMC48107.2020.9148521

[6] Banos, Oresti, Rafael Garcia, Juan A. Holgado-Terriza, Miguel Damas, Hector Pomares, Ignacio Rojas, Alejandro Saez, and Claudia Villalonga. "mHealthDroid: a novel framework for agile development of mobile health applications." In *Ambient Assisted Living and Daily Activities: 6th International Work-Conference, IWAAL 2014, Belfast, UK, December 2-5, 2014. Proceedings 6*, pp. 91-98. Springer International Publishing, 2014. https://doi.org/10.1007/978-3-319-13105-4_14

[7] Banos, Oresti, Claudia Villalonga, Rafael Garcia, Alejandro Saez, Miguel Damas, Juan A. Holgado-Terriza, Sungyong Lee, Hector Pomares, and Ignacio Rojas. "Design, implementation and validation of a novel open framework for agile development of mobile health applications." *Biomedical engineering online* 14 (2015): 1-20. https://doi.org/10.1186/1475-925X-14-S2-S6

[8] Bastos, Lucas, Thais Tavares, Denis Rosário, Eduardo Cerqueira, Aldri Santos, and Michele Nogueira. "Double authentication model based on ppg and ecg signals." In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 601-606. IEEE, 2020. https://doi.org/10.1109/IWCMC48107.2020.9148521

[9] Sundarasekar, Revathi, M. Thanjaivadivel, Gunasekaran Manogaran, Priyan Malarvizhi Kumar, R. Varatharajan, Naveen Chilamkurti, and Ching-Hsien Hsu. "Internet of things with maximal overlap discrete wavelet transform for remote health monitoring of abnormal ECG signals." *Journal of medical systems* 42 (2018): 1-13. https://doi.org/10.1007/s10916-018-1093-4

[10] Filip Wasilewski, "Wavelet daubechies 4 (db4) properties." Wavelet Browser by PyWavelet (2022).

[11] Pan, Jiapu, and Willis J. Tompkins. "A real-time QRS detection algorithm." *IEEE transactions on biomedical engineering* 3 (1985): 230-236. https://doi.org/10.1109/TBME.1985.325532

[12] Suraj Donthi. "Support Vector Machines, Dual Formulation, Quadratic Programming & Sequential Minimal Optimization." *Towards Data Science* (2021).

[13] Perez-Siguas, Rosa, Hernan Matta-Solis, Eduardo Matta-Solis, Luis Perez-Siguas, Hernan Matta-Perez, and Alejandro Cruzata-Martinez. "Emotion Analysis for Online Patient Care using Machine Learning." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 30, no. 2 (2023): 314-320. https://doi.org/10.37934/araset.30.2.314320

[14] Milecia McGregor. "SVM machine learning tutorial – what is the support vector machine algorithm, explained with code examples." *Free Code Camp* (2021).

[15] Rohith Gandhi. "Support vector machine — introduction to machine learning algorithms." *Towards Data Science* (2018).

[16] Bruno Stecanella, "Support vector machines (SVM) algorithm explained." Monkey Learn Blog (2017).

[17] Majid, Hanafi, Syahid Anuar, and Noor Hafizah Hassan. "TPOT-MTR: A Multiple Target Regression Based on Genetic Algorithm of Automated Machine Learning Systems." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 30, no. 3 (2023): 104-126. https://doi.org/10.37934/araset.30.3.104126

[18] MathWorks. "Remove trends from data." *The Mathworks Inc.* (2022).

[19] D'Aloia, Matteo, Annalisa Longo, and Maria Rizzi. "Noisy ECG signal analysis for automatic peak detection." *Information* 10, no. 2 (2019): 35. https://doi.org/10.3390/info10020035

[20] Jain, Anil K., Arun Ross, and Salil Prabhakar. "An introduction to biometric recognition." *IEEE Transactions on circuits and systems for video technology* 14, no. 1 (2004): 4-20. https://doi.org/10.1109/TCSVT.2003.818349

[21] Math Courses." Mean vs. Median | Usage, Calculation & Examples." *study.com* (2021)

[22] Akash Takyar. "A comprehensive exploration of various machine learning technique." *LeewayHertz*.

[23] Ray Hsu. "SVM: Why maximize margin." *Ray's Data Science Home* (2022)

[24] Yang, Jiawei, Zeping Wu, Ke Peng, Patrick N. Okolo, Weihua Zhang, Hailong Zhao, and Jingbo Sun. "Parameter selection of Gaussian kernel SVM based on local density of training set." *Inverse Problems in Science and Engineering* 29, no. 4 (2021): 536-548. https://doi.org/10.1080/17415977.2020.1797716

[25] Jarchi, Delaram, and Alexander J. Casson. "Description of a Database Containing Wrist PPG Signals Recorded during Physical Exercise with Both Accelerometer and Gyroscope Measures of Motion." *Data* 2, no. 1 (2017). https://doi.org/10.3390/data2010001

[26] MathWorks. "Assess classifier performance in classification learner." *The MathWorks Inc.*(2022).

[27] Yang, Jiawei, Zeping Wu, Ke Peng, Patrick N. Okolo, Weihua Zhang, Hailong Zhao, and Jingbo Sun. "Parameter selection of Gaussian kernel SVM based on local density of training set." *Inverse Problems in Science and Engineering* 29, no. 4 (2021): 536-548. https://doi.org/10.1080/17415977.2020.1797716