Received: 22 September 2022 | Revised: 25 October 2022 | Accepted: 6 December 2022

IET Biometrics

DOI: 10.1049/bme2.12105

ORIGINAL RESEARCH

Activity-based electrocardiogram biometric verification using wearable devices

Hazal Su Bıçakcı | Marco Santopietro | Richard Guest

School of Engineering, University of Kent, Canterbury, UK

Correspondence

Hazal Su Bıçakcı, Marco Santopietro and Richard Guest, School of Engineering, University of Kent, Jennison Building, Canterbury, UK.
Email: hb488@kent.ac.uk; ms2101@kent.ac.uk and r.m.guest@kent.ac.uk

Funding information

University of Kent

Abstract

Activity classification and biometric authentication have become synonymous with wearable technologies such as smartwatches and trackers. Although great efforts have been made to develop electrocardiogram (ECG)-based biometric verification and identification modalities using data from these devices, in this paper, we explore the use of adaptive techniques based on prior activity classification in an attempt to enhance biometric performance. In doing so, we also compare two waveform similarity distances to provide features for classification. Two public datasets which were collected from medical and wearable devices provide a cross-device comparison. Our results show that our method is able to be used for both wearable and medical devices in activity classification and biometric verification cases. This study is the first study which uses only ECG signals for both activity classification and biometric verification purposes.

KEYWORDS

activity classification, behavioural biometrics, biometrics (access control), ECG biometrics, emotion recognition, wearable devices

1 | INTRODUCTION

Biometric authentication has become commonplace within technological solutions coupled with the rise of security demands in recent years. Electrocardiogram (ECG) is the measurement of the heart's electrical activity utilising electrodes connected to the human body. Every living creature has a unique heartbeat rhythm which is not easily reproducible. For this reason, ECG provides a useful candidate for authentication systems [1–4]. Electrocardiogram can be collected by wearable and medical devices with different electrode types and different numbers of electrodes (typically 1, 2, 3, 6, or 12 leads). Moreover, the locations of the electrodes are important for ECG recordings using devices such as chest bands [5, 6], armbands [7], wrist bands [8], and Holter devices [9] that measure ECG from different body parts.

A number of key features extracted from to ECG waveform are used to understand heart conditions and ECG-based applications. Each heartbeat consists of *P*, *Q*, *R*, *S* and *T* complexes. An ECG heartbeat is represented in Figure 1 [10]. In healthy individuals, *P*, *R* and *T* complexes have positive peaks, while *Q* and *S* complexes have negative troughs. As a general trend, the time interval between complexes is generally consistent within particular healthy subjects, however, distorted peaks and unbalanced time intervals show heart irregularity.

There are 3 phases in a cardiac cycle. The first is an atrial depolarisation known as the *P* wave, the first positive deviation of the ECG. The second is the QRS complex which includes *Q*, *R* and *S* waves. After atrial depolarisation, the signal passes the Hiss-Purkinje systems and ventricular depolarisation was activated. In the third phase, the *T* wave is represented with ventricular repolarisation. Cardiac cycles repeat with regularity unless there is another effect (physical activity, medical drug interventions, heart failure etc.).

Any biometric ECG-based system must be invariant or aware of different activity situations to ensure consistent performance. These activity situations might be emotional, such as stress and excitement, or physical such as running, walking and sitting. Emotions occur spontaneously without any conscious effort. Emotions are accompanied by physiological changes, and this can be described as a mental state [11, 12]. Although

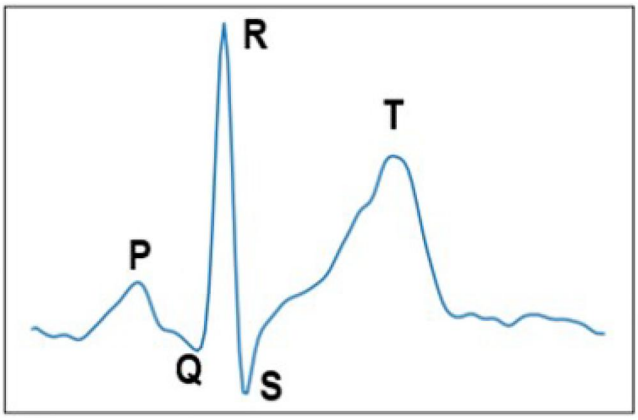

FIGURE 1 A single heartbeat electrocardiogram (ECG) signal with P wave, QRS complex and T wave representation [10].

emotions are subjective, common emotions such as being angry, sad and happy can be classified.

Our aim in this study is to utilise physical activity and emotion detection to adaptively select biometric authentication mechanisms to enhance overall performance using only ECG recordings. *Q*, *R*, *S* and *T* points were used as reference points for Manhattan and Euclidean distances between these points and subsequently used as features. We detect and utilise ground-truthed emotional and physical activity states within pre-collected data to assess tuned biometric authentication models. Data were assessed across both medically approved and wearable ECG recorders thereby allowing an assessment of device stability. Using only ECG recordings and new features for biometric recognition and activity classification is the contribution of this study to science.

2 | RELATED WORKS

Related works to this study are introduced in this Section as ECG biometric verification and identification, physical activity classification and activity-aware biometric systems, and emotion classification and distance metrics comparisons.

2.1 | Electrocardiogram biometric verification and identification

Medical devices have typically 6 or 12 wet type leads with a high sampling frequency (generally above 1 kHz), while wearable devices have 1 or 3 dry type sensors with a relatively low sampling frequency (under 1 kHz). The placements of attached sensors, the number of sensors and the type of sensors may affect ECG recordings. For these reasons, medical ECG recording systems give more reliable and less noisy signals than wearable ECG recorders [13]. Therefore, many studies have been conducted to understand different device performances with respect to biometric authentication such as Ingale et al. [14] who studied on comparative analysis of ECG-based biometric authentication with 6 different ECG datasets representing different devices. They used an extended Kalman filter and an Infinite Impulse Response (IIR) filter with ECG-beat and RR interval segmentations. They extracted 30 fiducial and non-fiducial features from ECG signals and they identified Equal Error Rate (EER) between 0.5% and 7% from fiducial features and a wider range of EERs from non-fiducial features. In the biometric verification process, Euclidean distance and Dynamic Time Warping methods were performed as a matching algorithm. While ECG-ID which consists of single lead healthy ECG recordings and MIT-BIH dataset which includes medical ECG recordings from unhealthy and healthy subjects have higher EER, other datasets which were collected by Biopac MP36 2 channel ECG recorder, 16-lead medical device and 2-lead palm and finger ECG sensors have relatively lower EER. They recognised that fixed window data segmentation, fiducial features and IIR filter gave better results than unfiducial features, Kalman filter and RR interval segmentation methods. Even if healthy and unhealthy subjects are considered in that study, there is no activity case or different simultaneously recorded devices.

ECG-based biometrics typically use fiducial and non-fiducial approaches for feature extractions [15]. While the fiducial approach is interested in time intervals between PQRST complexes and their amplitudes, the non-fiducial approach is related to the transform domain and model-based features. Choi et al. [16] acquired 60-s signals from 127 subjects with a 2-lead mobile sensor. Eight fiducial features were extracted and were classified using Support Vector Machine (SVM), Simple Logistic, Naïve Bayes (NB), Random Forest (RF), Adaboost (AB), Multi-Layer Perceptron (MLP), Bayes Net and radial basis function Kernel-based SVM architectures for single-beat and multi-beat biometric authentication scenarios. In the single-beat verification scenario, they identified between 4.46% and 9.51% EER from different classifiers. While SVM showed the best performance, the worst result came from NB. However, biometric authentication needs more beats to achieve more reliable results. For this reason, they tested their systems from 3 to 15 s of data using the best-performing SVM classifier. System EERs decreased to 8% (3 s) to 1.87% (15 s) EER. These results show that when testing time increases, EER decreases [16]. Krasteva et al. [17] compared 12-leads ECG signals from 460 patients in a verification configuration. They used 8 fiducial features for each of the 12-leads and found that they obtained their best results from the frontal plane leads (I, -aVR or II). Half of the data were used for training with equal sizes of genuine and imposters while another half of the data were used for testing with more imposters samples using the Linear Discriminant Analysis (LDA) classifier for biometric verification. They obtained several EERs according to different ECG-leads configurations from 3.7% to 32.4% EER. Pavia et al. [18] used a single-beat ECG dataset for extracting fiducial points in identification. They compare different training durations with an SVM classifier, and they found the best identification accuracy as 97.5%. Sidek et al. [19] used ECG signals with different activities in their fiducial feature-based person identification study. They used normalised QRS complex features with a MLP classifier and they achieved 96.1% classification

accuracy. Kim et al. [20] and Lehmann et al. [21] recognised that ECG recordings which were collected on different days vary because of different activity and emotional conditions. In ref. [20], they presented a biometric authentication method for short ECG recordings. They used QT interval correction with Sum of Squared Difference (SSD), Gated Recurrent Unit (GRU) and Convolutional Neural Network (CNN) deep learning (DL) methods. In one case, they used 2 different days of data for both training (1 record of 83 subjects and 25 subjects per day) and testing (1 record of 83 subjects and 25 subjects per day and 4-fold cross-validation). When they selected 1 pulse for training, they obtained 13.58%, 1.85%, and 2.25% EERs and when 3 pulses were used in training, 9.78%, 2.82%, and 2.29% EERs for SSD, GRU, and CNN method respectively in both cases. In another case, they used 1 record of 83 subjects in training and 2 records of 25 subjects with 2-fold cross-validation in testing data from different days. 12.44%, 2.15%, and 3.81% EERs for 1 pulse and 9.43%, 3.66%, and 3.1% EERs for 3 pulses were obtained by SSD, GRU and CNN methods respectively. Even if they obtained prominent results, the difference between sensitivity and specificity is quite large. This shows that genuine and imposter samples may not be equal. In Ref. [21], they used a chest band for ECG recordings with 256 Hz from 20 subjects during a week. They used peak-to-peak intervals of QR, RS, and QS as features with RF, neural network (NN) and SVM classifiers. 2000 samples in training and 500 samples in testing for both genuine and imposters were used in their study. According to data which were collected from different days scenarios, EERs are between 19.54% and 21.91% for RF, between 20.16% and 26.17% for NN and from 26.77% to 28.08% for SVM classifiers. The best results were obtained when 3 days of data were used in training while the worst EERs were found when 1 day of data was used in training.

2.2 | Physical activity classification and activity-aware biometric systems

Activity recognition using mobile and wearable devices before the biometric user identification and verification were studied by Sriram et al. [22], Mekruksavanich et al. [23] and Batool et al. [24]. However, activity classification studies always focus on ECG recordings with additional sensors such as gyroscopes and accelerometers. Our study used only ECG signals for both activity classification and biometric verification. In ref. [22], they used ECG and accelerometer data to prove that an activity-aware system which means activity classification before biometric identification has better results than the activity-unaware system which means biometric identification without any activity classification. Electrocardiogram and accelerometer data to create activity-aware system features and ECG data for the activity-unaware system were used in their study. They used 4-s data windowing with k-Nearest Neighbour (KNN) and Bayesian Network (BN) for activity classification. They obtained 82.78% and 84.88% identification precision rates for KNN and BN respectively on their activity-aware system. Moreover, they obtained a maximum of 78.55% and 81.39% identification precision rates on the activity-unaware system. Apart from the activity-aware system, they also checked the biometric verification results to see their device performance for 17 subjects. They used 80 s of genuine data for verification testing with different imposter sizes. They obtained 15.52%, 15.29%, 15.91%, 15.25%, and 16.11% EERs using BN classifier combined ECG and accelerometer data for 3, 7, 8, 11, and 15 imposter samples respectively. In ref. [23], activity recognition was processed before biometric identification. Gyroscope and accelerometer sensor data were used to recognise 12 daily activities in USC HAD dataset and 6 activities in UCI HAR dataset with CNN and Long Short-Term Memory DL structures (LSTM). Activities were grouped as dynamic (i.e. walking, jumping, running etc.) and static activities (i.e. sitting, lying, standing etc.). The UCI HAR dataset has 30 subjects and the USC HAD has 14 subjects. Mean activity classification accuracy rates were 91.24% by ConvLSTM using the UCI HAR dataset and 87.77% by CNN-LSTM using USC HAD dataset. Dynamic activities (walking-related activities) were classified better than static activities (posture-related activities). The study obtained the best mean user identification rates with 91.77% from the UCI HAR dataset and 92.43% from USC HAD dataset using walking-related activity data. In ref. [24], gyroscope data from 19 subjects with an IoT device was used to recognise three daily activities (walking, sitting, and standing) and biometric identification separately. Their study aimed to prove that IoT data can be used with the RF classifier in activity classification and biometric identification. They obtained a 93% accuracy rate in activity recognition and from 81% to 100% accuracy rates in biometric user identification using 10-fold cross-validations. Another study by Li et al. [25] used ECG and accelerometer together for 2 daily physical activity detections with an SVM classifier. They obtained a 79.3% activity recognition accuracy rate using data which were collected from different days and a 97.3% activity recognition accuracy rate using data which were collected from the same day.

2.3 | Emotion classification and distance metrics comparisons

There are few studies that assess emotional state and activity classification using ECG. Selvaraj et al. [11] used ECG signals to classify six emotions: happiness, sadness, fear, surprise, disgust, and anger. Sixty healthy participants watched the video clips during the experiment and ECG signals were collected from the left and right hand and the left leg at the same time. Rescaled Range statistics, Finite Variance Scaling (FVS), a Wavelet Transform (WT) and Empirical Mode Decomposition which are known as non-linear approach features were used to obtain Hurst parameters. K-Nearest Neighbour, Regression Tree (RT), NB and Fuzzy KNN (FKNN) classifiers were used for emotion classification with 70% training and 30% testing data. They achieved the best performance with 92.87% classification accuracy using FVS features and the FKNN classifier. The

WeSAD (Wearable Stress and Affect Detection) dataset was used for emotion recognition and classification. Linear Discriminant Analysis, Decision Tree (DT), RF, AB, KNN [26], Multimodal-Multisensory Sequential Fusion (MMSF) [27] and Self-supervised representation learning [28] methods achieved 66.29%, 83%, and 95% emotion recognition accuracy rates respectively. In ref. [26], LDA achieved a higher accuracy rate of 66.29% on three-class classification (baseline, stress and amusement) using only ECG data. Lin et al. [27] used the MMSF model to classify three emotions and they compared different types of data influences on accuracy rates. According to classification accuracy rates, the highly influential signals are respiration, electromyography (EMG) and ECG signals respectively. They obtained 83% classification accuracy on three-class using all chest signals while using only ECG signals brings 78% classification accuracy. Sarkar et al. [28] used single-task CNN to classify 4 emotions (baseline, stress, meditation, and amusement) and they obtained 95% accuracy rates.

Template matching algorithms were compared in many studies [29, 30] such as a Manhattan distance used in skin texture-based personal identification [31] and arrhythmia classification using KNN [32]. Aggarwal et al. [33] proved that Manhattan distances are preferable to other metrics, especially in the high dimensional cases. Perpinan et al. [30] compared Manhattan distance and Euclidean distance metrics on metabolic syndrome classification using heart rate variability features with the k-means algorithm and they found Manhattan distance outperformed Euclidean distance. In Lee et al. [29], Manhattan distance-based matching algorithm outperformed Euclidean, cosine similarity and cross-correlations metrics in single beat ECG-based individual biometric identification. In 2 heartbeats and 3 heartbeats ECG cases, Euclidean distance metrics gave higher results.

The proposed study is the first to use Manhattan and Euclidean distances between QR, RS and QS points as feature sets. As a second point, even if there are many studies to compare different classifier performances on activity classification [22, 26], and biometric verification [16], there is no study on ECG-based activity classification before biometric verification. This study first classifies which activities the data come from. Then the classified data is then used in the biometric verification. Created features were used for activity classification with KNN, SVM, DT, Bagged Tree Ensemble (BT), Subspace K-NN Ensemble (S-KNN) and NN classifiers. According to activity classification accuracy rates, classifiers are compared using data from two publicly available datasets. Emotional (neutral, stress, and amusement states) and physical activity (standing, walking, and resting) data are classified and evaluated separately. In this way, the performance of generated features with different classifiers is compared. Subspace K-NN Ensemble classifier has the highest classification rates in all activity classification cases. In the proposed method, Data classified by S-KNN according to their activities go to the biometric verification. In biometric verification, data from each activity are performed by NB and DT classifiers to obtain binary decisions (genuine or imposter).

3 | MATERIALS AND METHODS

A hardware experimental setup was run on an Intel(R) Core (TM) i7-3537U CPU and 8 GB RAM computer using MATLAB 2021a. A classical machine learning model with new features was used to explore activity classification effects on biometric verification. Prior to exploring activity classification, we investigated the performance of a novel biometric verification system. In a second set of experiments, activity classification was performed to classify physical activities and emotional activities using these data to select optimised biometric classifiers. All data used in this study come from WeSAD [26] and Vollmer [34] databases. No previous study using these datasets include about these datasets includes biometric verification performance comparisons after activity classification. In this section, datasets, models and experiment protocols are explained.

3.1 | Databases

The WeSAD dataset consists of physiological data (ECG, EMG, respiration etc.) which were collected from chest-worn RespiBAN and wrist-worn Empatica E4 devices. Fifteen healthy subjects (3 females and 12 males) were analysed from this dataset. Data at a 700 Hz sampling frequency with 16-bit analog-to-digital converter resolution was obtained from the RespiBAN device [26]. Initially, data were collected at a baseline condition where subjects can sit, stand at a table and read magazines for 20 min. This condition was collected to see the neutral emotional state of subjects. In the amusement condition, 11 funny videos which were selected from the corpus [35] were watched by participants for approximately 6 min. After each video, there was a neutral sequence for 5 s. In the meditation condition, subjects were in a sitting position and tried to return to a neutral emotional state for 7 min. In the stress condition, subjects experienced a public speaking and a mental arithmetic task for 5 min each task. This study focussed on data collected during the baseline, stress and amusement periods. The quantity of data from each activity must be equal to create robust activity classification, therefore 6-min data from each activity (a total of 18 min) from the RespiBAN device were used in this study.

The Vollmer dataset [34] consists of 13 healthy subjects' data collected simultaneously with 5 different devices. These devices are the clinically certified (i.e. medical-based) NeXus-10 MKII (8000 Hz) [36], eMotion Faros [37], SOMNOtouch NIBP (512 Hz) [38], and the consumer products Hexoskin Hx1 (256 Hz) [39] and Polar RS800 Multi (1000 Hz) [5]. Only the Polar device cannot measure raw ECG data. However, R-peaks were measured by the Polar device which were selected as reference points. According to these reference points, other R-peaks which were measured by other devices were synchronised by Vollmer's study [40]. The reason for synchronising the positions of the R peaks measured from each device is to equalise all devices at a sampling frequency of

256 Hz and all heartbeat locations are the same. They provided synchronised signals in PhysioNet [41].

Data were collected during 4 tasks: resting, walking on the treadmill at 1.2 m/s speed, standing still position and uphill walking on the treadmill at the same speed with 15% track inclination. Each task has a 5-min recording time separately. The sensor placements of the Vollmer dataset are shown in Table 1.

These two datasets were chosen due to limited prior experimentation for the biometric verification and activity effects on biometric performances using only ECG data. In addition, the chance to compare device performances played a vital role in choosing Vollmer's dataset. The WeSAD dataset was selected to explore new features' effects on emotional activities.

3.2 | Pre-processing

In the Vollmer dataset, there are several minutes of recorded data not related to the experimental setup. These recorded data before the tasks were used to analyse the baseline and evaluate the stability of the ECG recording to 256 Hz sampling frequency by Vollmer [34, 40]. In our study, 20-min recordings per person (5-min per activity) were selected using provided data labels in the dataset. The sampling frequency of each device was reduced from 256 to 200 Hz for easy calculation of PQRST points and filtering/smoothing signals. Low-frequency noise such as baseline drift, high-frequency noises and power-line interference components generally exist in ECG signals. These noises can be produced by muscles, external interferences, activities and baseline wander noise. In the pre-processing phase, the signal was filtered by a band-pass filter. A 0.5 Hz low cut-off frequency filter was used to remove baseline wander and a 45 Hz high cut-off frequency with three ordered band-pass filters was used in pre-processing. In addition, a mean filter was used for signal smoothing.

In the WeSAD dataset, recording durations were variable for each person and each activity. Subjects were randomly divided into two groups, each with a different order of activities. The activity-less signals between tasks were removed from the recordings and each activity task was manually equalised at 360-s per activity. Sampling frequencies were reduced from 700 to 200 Hz and all types of noises were filtered by the same filtering methods in the Vollmer dataset. The sampling frequency was reduced to 200 Hz to equalise conditions for both datasets.

In a realistic biometric verification scenario, the recording time is generally narrow. Therefore, the selection of the most appropriate time window to sample the signal is extremely important. Abdul-Kadir et al. [42] compared different time window sizes to explore an optimum for ECG atrial fibrillation (AF) recognition and they found that a window duration of 4 s was the most suitable for their features based on Artificial NN and SVM classifiers' results (95.3% and 95% AF recognition accuracy rates respectively). Inspired by this study, 4-s time windows and a 1-s delay between windows were used in the pre-processing. The delay was selected to avoid overfitting and to have a clearer separation between windows. In this way, the PQRST complexes of all windows are completely separated without any loss complexes.

3.3 | Feature extraction

Twenty eight fiducial features *(Qamp, Ramp, Samp, Tamp, QQman, RRman, SSman, TTman, QQeuc, RReuc, SSeuc, TTeuc, Manhattan, and Euclidean distances of QR, ST, RS, RT, QS, QT, RST, and QRS)* were extracted from each sample in the preprocessed matrix. Maximum values of each peak were named as *X*. Manhattan distances between two peaks' maximum values were shown as *(XX)man*. Euclidean distances between two peaks' maximum values were shown as *(XX)euc*. Pan-Tompkins algorithm was used for finding *R* peak positions. Maximum (to find *P* and *T*) and minimum (to find Q and S) values in a specific confident interval of milliseconds before or after each *R* peak. An explanation of all features was shown in Table 2.

(XX)man features were found using Manhattan distances between two peak values. It can be explained by Equation 1:

$$|(x_{i+1}) - (x_i)| \tag{1}$$

(XX)euc features were calculated with Euclidean distances between two peaks. It can be explained by Equation 2:

$$\sqrt{((x_{i+1}) - (x_i))^2} \tag{2}$$

where "x" is a selected peak (*Q*, *R*, *S* or *T*) and "i" is the location of the selected peak. To calculate XYZman ($Q = X$,

TABLE 1 Sensor placements of Vollmer's dataset

Device Name	Number of Sensors	Locations of ECG sensors	# of ECG channels	Sampling Frequency	ECG Electrode Type
SomnoTouch	6	Left chest, left & right abdomens	3	512 Hz	Attachable patches
Emotion Faros 360	3	Left right chest, right abdomen	3	1000 Hz	Attachable patches
Nexus-10 MKII	4	Left right chest, right abdomen	3	8000 Hz	Attachable patches
Hexoskin Hx1	4	Top left right abdomens, bottom left abdomens	3	256 Hz	Textile
Polar RS800 multi	1	Top centre abdomen	1	1000 Hz	Bipolar chest strap

TABLE 2 Explanations of features are represented as X, Y and Z symbols

Features	Values of X,Y and Z	Description	
X	Q_{amp}, R_{amp}, S_{amp}, T_{amp}	$X = \{x_1, x_2, \ldots, x_n\}$	Peak amplitudes
XXman	QQ, RR, SS, TT	$d(X_{i+1}, X_i) = \{\lvert X_{i+1} - X_i\rvert\}_n$	
XYman	QR, ST, RS, RT, QS, QT	$d(X_i, Y_i) = \{\lvert X_i - Y_i\rvert\}_n$	Manhattan distances
XYZman	QRS, RST	$d(X_i, Y_i, Z_i) = \{\lvert X_i - Y_i\rvert + \lvert X_i - Z_i\rvert + \lvert Y_i - Z_i\rvert\}_n$	
XXeuc	QQ, RR, SS, TT	$d(X_{i+1}, X_i) = \left\{\sqrt{(X_{i+1} - X_i)^2}\right\}_n$	Euclidean distances
XYeuc	QR, ST, RS, RT, QS, QT	$d(X_i, Y_i) = \left\{\sqrt{(X_i - Y_i)^2}\right\}_n$	
XYZeuc	QRS, RST	$d(X_i, Y_i, Z_i) = \left\{\sqrt{(X_i - Y_i)^2 + (X_i - Z_i)^2 + (Y_i - Z_i)^2}\right\}_n$	

$R = Y, S = Z$), $|X - Y| + |X - Z| + |Y - Z|$ formula for all QRS samples were used step-by-step. There was the same process for RSTman and Euclidean distance was used for XYZeuc and RSTeuc cases.

To select the most powerful feature sets, all features were examined separately according to their direct biometric verification performances. According to these examinations, only 1-peak-based features (XXman, XXeuc) have the lowest performance than 2-peak (XYman, XYeuc etc.) and 3-peak (XYZman, XYZeuc etc.) features. Although there were close results between 2-peak features and 3-peak features, 2-peak features have better performances during low training cases. Finding T waves' amplitudes might be problematic, especially in the activity case. In all cases which included T wave amplitudes, biometric verification performances were more insufficient than others. For this reason, features with T waves in 2-peak features were not calculated. According to all experiments, feature sets which have higher biometric performances were selected as *QRman, RSman, QSman, QReuc, RSeuc, and QSeuc*.

The absolute value of subtraction is always positive and in the Manhattan distance formula, the absolute value of the subtraction is calculated. For this reason, we can say $|X - Y| = |Y - X|$. The square of the parenthesis of subtraction is also always positive and in the Euclidean distance formula, the square of the parenthesis of subtractions is used so it can be explained by Equation 3:

$$\sqrt{(X - Y)^2} = \sqrt{(Y - X)^2} \tag{3}$$

According to these situations, there is no need to try the RQman and RQeuc while the QRman and QReuc are being calculated.

Previous studies used fiducial features such as time intervals between two peaks (XX and XY interval) and peaks and deep values of PQRST complexes and non-fiducial features such as WT. This study introduces new fiducial features using Manhattan and Euclidean distances of Q, R, and S amplitude values.

3.4 | Classification models

This study explored the performance of different features, classification models and data test set sizes for five different ECG recorders. There are two different verification pipelines considered in this study: one utilising activity-aware biometric verification (System B) and the other direct biometric verification without prior activity classification (System A). K-Nearest Neighbour, SVM, DT, BT, S-KNN and NN classifiers were explored for activity classification, whilst NB and DT were used in the biometric verification sections. Six classifiers were used to evaluate the performance of the generated features in different classifiers. The success of the created features in activity classification has been observed. Then, based on the classifiers that are frequently used in ECG-based studies [43–45], the number of classifiers for biometric verification was reduced to two, namely NB and DT.

These classifiers are commonly used for ECG-based machine learning models [46]. There are many differences between the classifiers. For instance, whilst all of them can be used for supervised learning, DT, KNN and NN are non-parametric methods, in other words, they work on non-linear data while NB can be used for linear data. Naïve Bayes produces probability estimations, whereas DT and SVM discriminate data. Even though DT is more flexible and easy to use, it can neglect some important values in training data; a situation that might cause accuracy losses. If there is adequate training data, NN gives better performances than DT for binary classification problems [47]. However, if we need an explanation of the decision-making process and multi-class classifications, DT is the better option. Support Vector Machine uses Kernel structures to deal with non-linear problems, however, it has an extremely long training time for large datasets. Some classifiers, such as DT, can be biased on certain features or the most representative data. This situation might result in lower classification accuracy. To solve this problem, ensemble modules are combined with basic classifiers.

According to Gul et al. [48], ensemble modules help less representative or non-informative data in classification. They used a large amount of data from 23 datasets to compare different ensemble structures with KNN, RF and SVM. They

added the non-informative features selected by Principal Component Analysis to the representative features. They observed S-KNN has better performance than SVM and RF for the classification of non-informative features. In our study, we used S-KNN and BT ensemble modules. In S-KNN, the function selects a number of elements, which is called a subspace, randomly from the feature vector in training. K-Nearest Neighbour classifier predicts the class labels for each selected subspace. After all subspace predictions, the majority vote of that predictions is calculated to achieve the final class label. In BT, many subsets of data from the training samples which are selected randomly with replacements are created. Each subset is trained by different DTs (in this study 100 DTs) to obtain predicted class labels. An average of each subset decision is calculated to achieve the labels of the final class.

This study represents new features and compares their efficiencies in activity classification and biometric verification scenarios. For this reason, using several classifiers gives us a wide perspective on different training and testing size results. The outline of the System A is shown in Figure 2 and the proposed System B is shown in Figure 3.

4 | EXPERIMENTS AND RESULTS

Several experiments were performed with different classifiers, three different training and testing sizes (50%, 40%, and 30% genuine samples in the System A. 80%, 60%, and 50% training samples in activity classification and 50% genuine samples in the biometric verification in the System B), two types of activity (emotional and physical activities) and five devices. Results are presented for biometric verification without activity classification, overall activity classification accuracy and activity-aware biometric verification subsections.

4.1 | System A

The data from each device has been tested separately. In each device data were split into two parts: training and testing. For training data, the data of 12 subjects from WeSAD dataset and 10 subject from Vollmer's dataset were selected. For testing data, unseen 3 subjects' data were separated. In each running process, selected unseen subjects were different and all subjects were used in the testing group at least once (i.e. selected subjects in testing were changed like person 1-2-3, 4-5-6 etc.).

According to the number of samples per person, genuine samples were calculated. For instance, if there are 100 samples per person, 30 samples selected from the testing subjects and 70 samples selected from the training subjects will be used for biometric verification. In this study, this situation was called as M-30% for Manhattan distance based features and E−30% for Euclidean distance based features. Imposter samples were randomly selected from the remaining unused samples. To compare results 40% and 50% of genuine sample sizes in testing set were used for both Manhattan and Euclidean distances features (i.e. M-40%, M-50%, E−40%, E−50%). The number of genuine samples and imposter samples are equal for both training and testing data. The number of each subject's genuine samples for training and testing is shown in Table 3. The number of genuine samples per activity were approximately equal. The number of subjects per dataset, the number of samples per device, activities and data collection procedures in each dataset were different. To achieve comparable results, we used the same ratio of genuine samples for all datasets and devices.

The most commonly used performance metrics of biometric systems are False Accept Rates (FAR), False Rejection Rates (FRR) and EER. False Accept Rates represents the percentage of accepted imposter samples in the system and FRR is the percentage of rejected genuine samples by the system. Equal Error Rate can be calculated from the intersection point of FAR and FRR.

As a first step, the biometric verification method was used to enable an understanding the new feature set performance. In this step, there is no activity classification system. For this reason, genuine and imposter samples were selected randomly according to their 30%, 40%, and 50% ratios. The RespiBan device from the WeSAD dataset and Faros, Somnotouch, Nexus-10 and Hexoskin devices from the Vollmer dataset were tested by NB and DT classifiers. Naïve Bayes classifier performances in Figure 4 and DT classifier performances in Figure 5 are shown below.

According to results in Figure 4, Manhattan distance-based features with the 30% of unseen participants' samples in testing

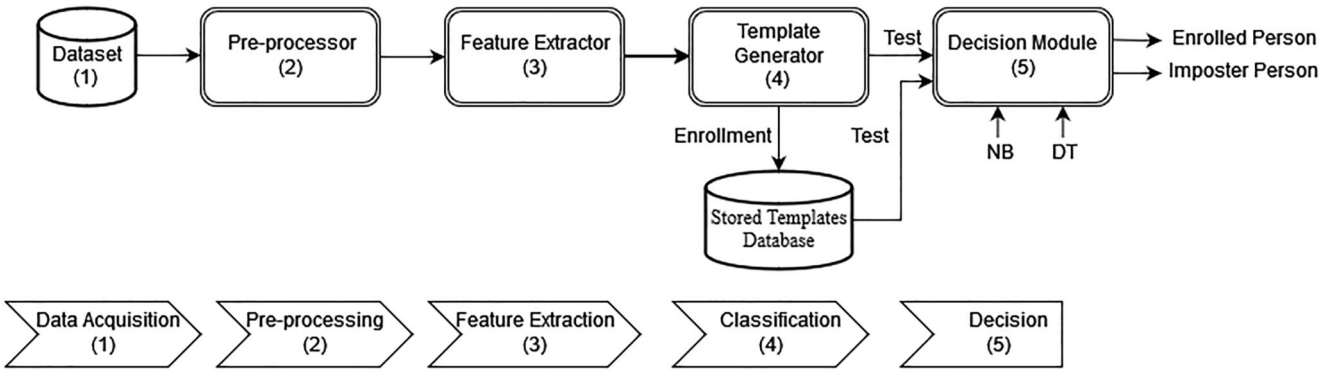

FIGURE 2 System A: Biometric verification model is shown at the top and the classification steps for biometric verification are shown below

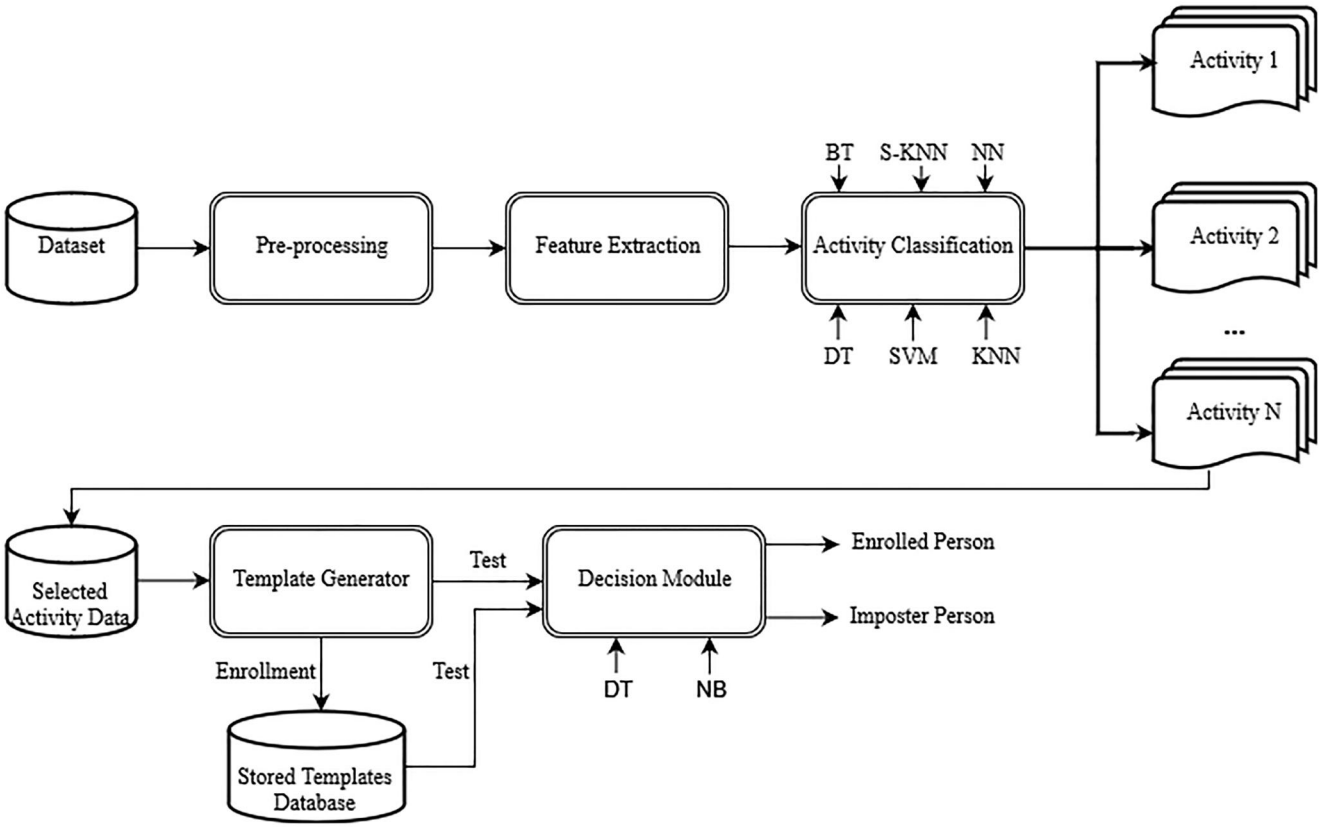

FIGURE 3 System B is an activity-aware biometric verification system

TABLE 3 The number of genuine samples in testing and training per person

Genuine samples		RespiBAN	Faros	SomnoTouch	Nexus-10	Hexoskin
50%	Train	534	731	770	737	837
	Test	534	731	770	737	837
40%	Train	641	878	924	876	1005
	Test	427	584	616	589	669
30%	Train	748	1024	1078	1032	1172
	Test	320	438	462	442	502

have the lowest EER (i.e. M-30%). For this reason, we would like to show the distribution of FAR and FRR with this testing set size. The RespiBAN device has 20.03% FAR and 19.59% FRR, the Faros device has 17.67% FAR and 17.05% FRR, the Somnotouch has 20.87% FAR and 19.21% FRR, the Nexus has 18.53% FAR and 18.71% FRR and the Hexoskin device has 17.69% FAR and 18.85% FRR in M-30% cases. In Euclidean distance-based features, there are the same trends as the corresponding Manhattan feature sets. While training set sizes increase, EER decrease and the lowest EER were obtained from E−30% cases.

In Figure 5, it is easily seen that from M-50% to M-30% as same as from E−50% to E−30%, EERs decrease for all devices. A large amount of testing samples causes higher EER. In practical use, more training samples yield more stable results. In our experiments, we proved that more training data presents low EERs.

As a general trend in Figures 4 and 5, Manhattan distance-based feature sets outperformed Euclidean features. Naïve Bayes classifier has lower EERs in all cases but NB and DT cases trends are same. Data which were collected from the RespiBAN and the SomnoTouch devices obtained higher EER. Faros, Nexus-10 and Hexoskin data have achieved similar performances. Even if the Hexoskin device is a consumer-based device, verification performances are similar with other medically-approved devices. The lowest EERs for each testing sizes (i.e. 50%, 40%, and 30%) were achieved from different devices. For instance, in Figure 5, while the lowest EER of the M-40% case was achieved from the Hexoskin device, the Nexus-10 device has better EERs in 30% and 40% sizes than other devices. In addition, in Figure 4, while the lowest EER of the E−30% case was achieved from the Faros device, the Nexus-10 device has lower EERs in E−50% and E−40%. This implies that device selection, feature selection and training/testing sizes affect biometric verification performance.

4.2 | System B, activity classification

To explore an improvement in biometric model performances, activity classification was used before biometric verification and we called this system as a System B. The

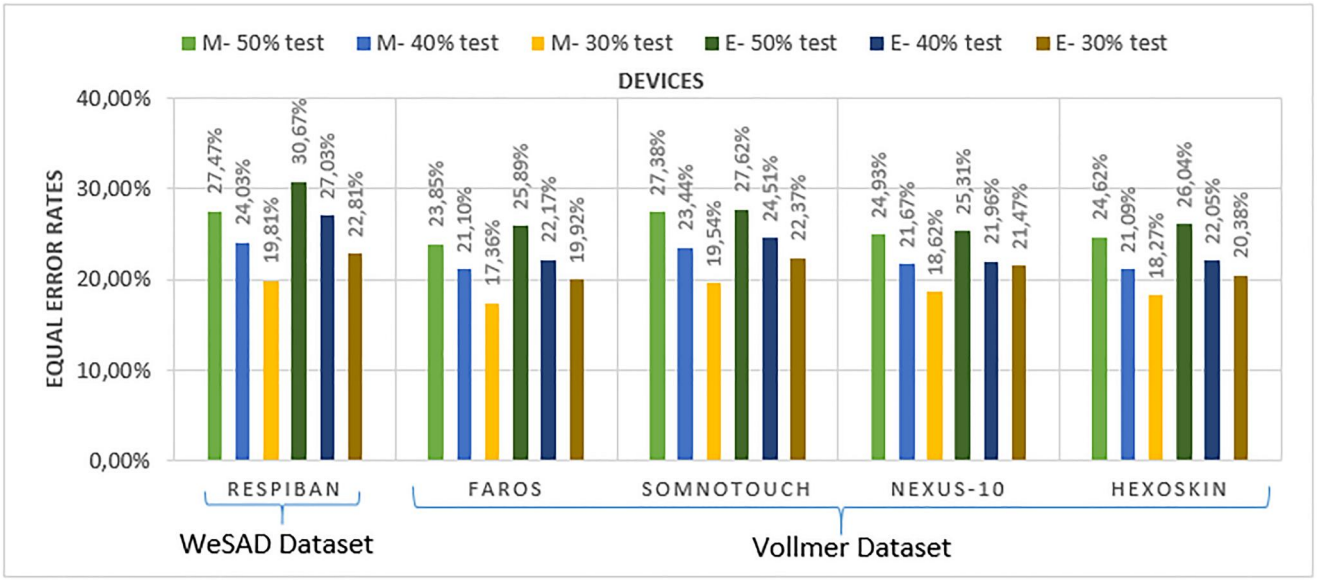

FIGURE 4 EERs in the System A were shown according to the Naïve Bayes (NB) classifier with 50%, 40%, and 30% genuine samples in test data. M is Manhattan distance-based features and E is Euclidean distance-based features

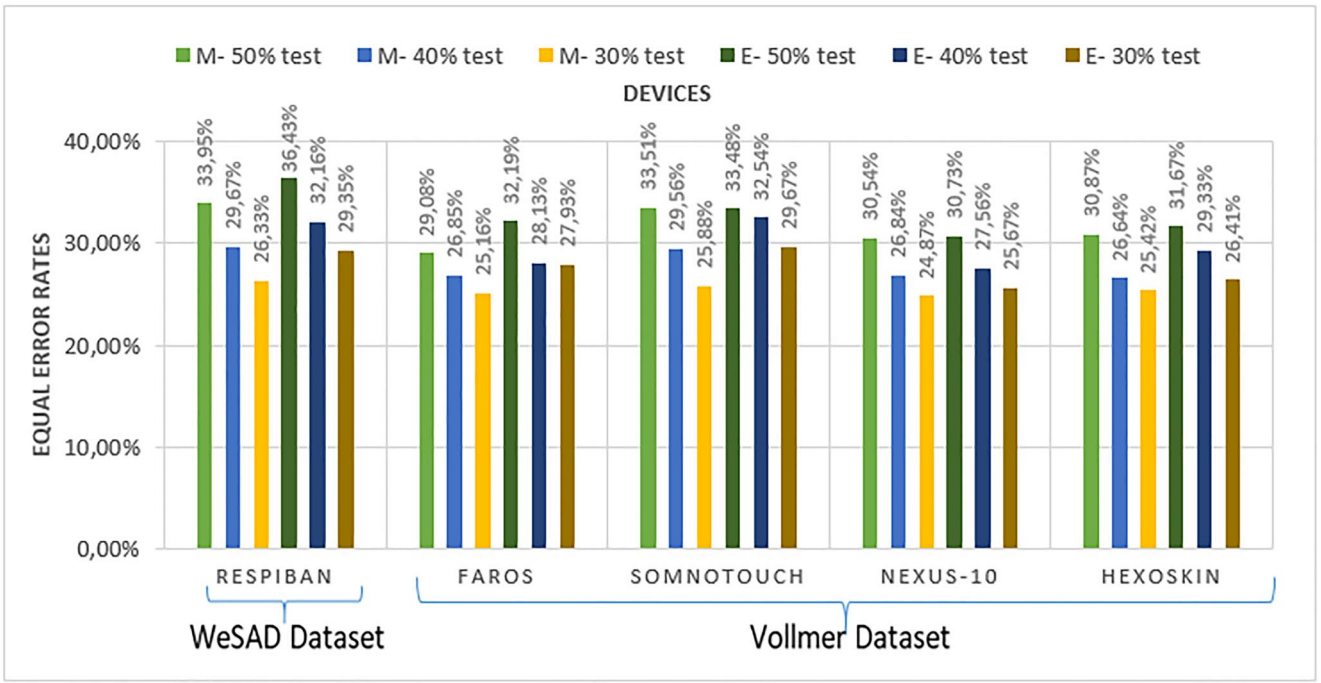

FIGURE 5 EERs in the System A were shown according to the Decision Tree (DT) classifier with 50%, 40%, and 30% genuine samples in test data. M is Manhattan distance-based features and E is Euclidean distance-based features.

main idea is to reduce the effects of the signal fluctuation that occurs when switching to a different activity. In system B, samples are first classified by their activity and subsequently set to tuned biometric verification stages, regardless of whether the activity is classified correctly or incorrectly. Inaccurate activity classification may affect biometric verification accuracy. However, using tuned biometric verification modules according to activity may increase overall system performance.

However, removing incorrectly classified samples will not be a realistic verification scenario. For this reason, incorrectly classified activity and emotion instances continued to be used with the labels of the classes to which they were assigned. A mean of the activity classification accuracy is shown in Table 4 for the Vollmer dataset with resting, walking, standing and uphill walking activities and Table 5 for the WeSAD dataset with baseline, stress and amusement activities. Decision Tree, SVM, KNN, BT, S-KNN, and NN classifiers were tested with

TABLE 4 Mean activity classification accuracy rates of the Vollmer dataset on different classifiers and feature sets

Vollmer dataset		Manhattan distance features				Euclidean distance features			
Device names		Faros	Somnotouch	Nexus	Hexoskin	Faros	SomnoTouch	Nexus	Hexoskin
DT	Tr-80%	60.30%	57.93%	53.80%	57.00%	58.00%	54.68%	52.11%	56.70%
	Tr-60%	61.40%	53.80%	52.20%	62.20%	60.10%	52.18%	52.60%	58.70%
	Tr-50%	51.80%	54.81%	53.00%	60.80%	60.61%	58.76%	52.41%	58.60%
SVM	Tr-80%	69.20%	70.46%	60.40%	74.80%	69.20%	68.32%	61.63%	76.02%
	Tr-60%	66.20%	70.01%	58.80%	74.50%	66.14%	68.10%	59.74%	72.50%
	Tr-50%	65.50%	66.83%	60.70%	74.50%	66.12%	66.93%	59.43%	74.10%
KNN	Tr-80%	66.00%	70.46%	58.50%	74.40%	65.10%	70.00%	61.18%	73.10%
	Tr-60%	62.30%	71.23%	59.40%	72.90%	63.10%	69.97%	58.12%	72.80%
	Tr-50%	62.70%	67.94%	56.30%	72.00%	62.30%	67.62%	55.67%	71.30%
BT	Tr-80%	88.20%	80.39%	84.90%	86.70%	86.20%	80.26%	86.14%	85.70%
	Tr-60%	85.80%	77.72%	81.90%	85.60%	89.10%	79.32%	82.50%	84.50%
	Tr-50%	86.60%	77.50%	80.80%	86.30%	87.30%	78.12%	83.24%	84.40%
S-KNN	Tr-80%	97.05%	96.88%	97.94%	96.94%	97.05%	96.56%	96.96%	96.26%
	Tr-60%	93.42%	93.58%	93.08%	94.77%	93.19%	93.48%	92.87%	94.19%
	Tr-50%	91.77%	91.64%	91.41%	92.78%	91.33%	91.35%	90.79%	92.35%
NN	Tr-80%	68.80%	66.00%	58.20%	67.30%	66.70%	65.61%	62.79%	67.20%
	Tr-60%	64.50%	62.95%	57.30%	70.00%	66.10%	64.53%	57.72%	69.80%
	Tr-50%	63.30%	64.00%	57.20%	68.40%	63.67%	63.86%	55.54%	67.90%

Note: Tr-80% represents 80% training, Tr-60% is 60% training and Tr-50% is 50% training samples.

different training and testing sizes. In Table 4 and Table 5, Tr-80% represents 80% of samples in the training set. Tr-60% has 60% of samples in training and Tr-50% is 50% of training samples.

It is easily seen that, as a general trend, from Tr-80% to Tr-50%, classification rates decrease. In the Vollmer dataset, both feature sets have similar results. The S-KNN ensemble classifier has higher activity classification accuracy rates than the other classifiers. For this reason, samples classified using S-KNN according to their activity were used during biometric verification with the NB and the DT.

As these datasets were collected in different circumstances, direct comparisons between them are not possible. If we compare the performance between devices, the Hexoskin and the Faros devices have better performances while the Nexus device has generally the worst performances for all classifiers. The Nexus device has the highest sampling frequency (8000 Hz). As the sampling frequency has been reduced to the lowest frequency of the Hexoskin device (256 Hz) in the Vollmer dataset, overall signal complexes may be preserved but distinctive features may be lost resulting in poor performance. In the WeSAD dataset, the Euclidean distance feature sets have higher accuracy rates using DT, SVM, KNN, NN classifiers. According to S-KNN and BT accuracy rates, Manhattan distance feature sets have higher results than Euclidean distance feature sets in both datasets.

4.3 | System B, biometric verification following activity classification

In Table 6, the performance of biometric verification following activity classification using Vollmer's dataset were presented in terms of EER for each activity case. EERs were presented according to their activities and the training/testing ratio of data in activity classification for each activity. The NB and the DT classifiers were used for biometric verification. Using the same procedure as described in the Section 4.1, classified samples were separated according to their activities.

Tested samples in Tr-80%, Tr-60%, and Tr-50% in activity classification pass through the biometric verification stage. The average EER of the Faros, Somnotouch, Nexus and Hexoskin devices were shown in Table 6 from the Vollmer dataset. The best performances are shaded according to each feature sets in the NB and the DT classifiers.

As a general trend, Manhattan distance-based feature sets outperformed Euclidean distance-based feature sets. The NB classifier has better performances than DT. If we look at the mean EER for each activity, EER in resting position are lower than other activities, while the EER in uphill walking is higher than others. The total number of samples in the verification stage decreases from Tr-50% to Tr-80%. In Table 7, biometric verification performances of the WeSAD dataset were presented in terms of EER for each activity case.

In the activity classification, incorrectly classified samples also passed to the biometric verification step with the incorrect class labels. For this reason, low activity classification accuracy rates in WeSAD dataset might cause higher EER. The best performances of Manhattan distance features come from the NB classifier in Tr-60%. False Accept Rates values of baseline, stress and amusement cases are 7.64%, 7.47%, and 7.61% respectively. False Rejection Rates values are 9.01% in baseline, 9.19% in stress and 9.06% in amusement situations.

On the other hand, the best performances of Euclidean distance features come from NB in Tr-50%. False Accept Rates values of baseline, stress and amusement activities are 8.86%, 8.65%, and 8.71% respectively while FRR values are 9.73%, 10.01%, and 9.98% respectively. There is no direct relationship between activities and EER so we cannot say that EER from specific activity are lower than other activities. In addition, NB has better performances than DT.

The speed of the recognition system was measured using CPU times in seconds. In direct verification cases, a timer was started at the beginning of the feature extraction and it was finished at the end of direct biometric verification without prior activity classification (i.e. System A). In the case of classification + verification, the timer starting point was the same as the direct verification case but it was stopped at the end of biometric verification with prior activity classification. In the case of only biometric verification after classification, the timer was started after the activity classification and it was finished at the end of biometric verification following activity classification. The recognition system speed was shown in Table 8.

TABLE 5 Mean activity classification accuracy rates of the WeSAD dataset on different classifiers and feature sets

WeSAD dataset	RespiBAN device	Manhattan	Euclidean
DT	Tr-80%	58.67%	58.98%
	Tr-60%	57.64%	56.13%
	Tr-50%	53.71%	55.39%
SVM	Tr-80%	65.86%	73.55%
	Tr-60%	63.00%	72.62%
	Tr-50%	62.34%	65.41%
KNN	Tr-80%	62.74%	68.51%
	Tr-60%	60.36%	67.83%
	Tr-50%	57.03%	66.79%
BT	Tr-80%	92.61%	82.46%
	Tr-60%	83.75%	81.92%
	Tr-50%	78.78%	77.76%
S-KNN	Tr-80%	93.67%	90.35%
	Tr-60%	86.46%	85.78%
	Tr-50%	82.10%	80.29%
NN	Tr-80%	70.52%	71.90%
	Tr-60%	68.56%	68.03%
	Tr-50%	65.45%	64.68%

Note: Tr-80% represents 80% training, Tr-60% is 60% training and Tr-50% is 50% training samples.

5 | CONCLUSIONS

For the activity classification, three cases were compared according to their training and testing sizes. Increased training data size have higher activity classification accuracies for System B and lower EERs for the System A in both feature sets. This study obtained that activity classification prior to biometric verification has better results than direct biometric verification. For this reason, activity classification is required

TABLE 6 Biometric verification performances in terms of average Equal Error Rate (EER) of all devices

	Feature Names		DT		NB	
			Manhattan	Euclidean	Manhattan	Euclidean
Activity names	Resting	Tr-80%	13.68%	13.71%	7.00%	7.45%
		Tr-60%	12.81%	13.51%	6.78%	7.31%
		Tr-50%	11.99%	12.61%	6.65%	7.07%
	Walking	Tr-80%	13.86%	13.56%	7.52%	7.73%
		Tr-60%	13.64%	13.40%	7.00%	7.61%
		Tr-50%	12.70%	12.85%	6.70%	6.99%
	Standing	Tr-80%	13.47%	13.83%	7.51%	7.63%
		Tr-60%	13.47%	13.88%	7.15%	7.48%
		Tr-50%	12.48%	13.01%	6.60%	7.25%
	Uphill walking	Tr-80%	13.69%	13.92%	7.64%	7.93%
		Tr-60%	12.52%	13.43%	7.04%	7.84%
		Tr-50%	12.40%	13.49%	6.88%	7.38%

Note: 50% of genuine samples were used in testing. The average value of all devices is represented.

TABLE 7 Biometric verification performances of RespiBAN device in terms of Equal Error Rate (EER)

	Feature Names		DT		NB	
			Manhattan	Euclidean	Manhattan	Euclidean
Activity names	Baseline	Tr-80%	15.08%	15.12%	8.32%	9.43%
		Tr-60%	15.01%	15.05%	8.32%	9.44%
		Tr-50%	14.88%	15.18%	8.48%	9.30%
	Stress	Tr-80%	15.01%	15.43%	8.41%	9.40%
		Tr-60%	15.06%	15.22%	8.33%	9.49%
		Tr-50%	15.20%	15.12%	8.34%	9.33%
	Amusement	Tr-80%	15.03%	15.18%	8.42%	9.43%
		Tr-60%	14.99%	15.02%	8.33%	9.51%
		Tr-50%	14.67%	15.04%	8.40%	9.35%

Note: 50% of genuine samples were used in testing.

TABLE 8 The recognition system speed in CPU execution time (seconds)

			Measured CPU Times (seconds)	
Dataset	Devices	Modality	Euclidean	Manhattan
WeSAD	RespiBan	Direct verification	0.8125	1.1875
		Classification + verification	2.7656	2.3438
		Only verification after classification	1.8281	1.2656
Vollmer	Hexos	Direct verification	0.9688	0.9531
		Classification + verification	3.6406	3.3281
		Only verification after classification	1.6965	2.0938
	Faros	Direct verification	0.9688	1.6723
		Classification + verification	3.9375	4.9219
		Only verification after classification	2.0469	2.8906
	Nexus	Direct verification	2.4844	1.9063
		Classification + verification	5.8438	5.9375
		Only verification after classification	2.3281	2.6719
	SomnoTouch	Direct verification	1.3750	0.9375
		Classification + verification	3.2656	3.1406
		Only verification after classification	1.9531	1.4063

before biometric verification. Medical and wearable devices had very close results in terms of activity classification and biometric verification.

The DT and NB classifiers were performed for both feature sets in biometric verification stages. Naïve Bayes generally gave lower EER than DT. As a general trend, when the number of testing samples are increased, EER also rose for System A. Conversely, in the System B, a larger amount of samples in testing led to lower EERs. System A samples have many variations between activities. Since signals that are similar to each other are gathered under the same activity group, there are relatively small variations in the System B.

To compare our results with other studies, we can cite examples from many studies. Many studies are concerned with activity recognition and classification with nonlinear [11] such as WT and linear approach features [25], such as skewness, kurtosis, mean of maxima and minima using different datasets. Moreover, deep fusion [27], CNN [28] and classical machine learning classifiers [26] were compared by many studies for emotion classification using WeSAD dataset. Even if refs. [11, 25, 26] and [27] used combined different sensor data such as EMG, accelerometer and ECG in their studies with mean, standard deviation, peak detection and heart rate variablity features, the proposed Manhattan and Euclidean distance-based features in our study has higher activity classification rates. Only CNN structure [28] with 95% accuracy rates in activity classification outperformed the proposed study with 93.67% activity classification accuracy rates.

When we examine the recognition speeds of the system, we see that the RespiBan device usually gives the lowest results in

seconds. This is because the WeSAD dataset consists of 15 min of data, while the Vollmer dataset consists of 20 min of data. By comparing Manhattan and Euclidean distances, we cannot say that one prevails over the other. Direct verification cases gave the shortest seconds in all cases while classification + verification (i.e. System B) cases gave the longest times in all cases. In the case of classification + verification, Hexos, Faros, and SomnoTouch devices achieved similar performances in the speed test. However, the Nexus device gave the longest recognition times, about 6-s.

In conclusion, the outcome of this work is ECG biometrics combined with activity classification based on ECG signal might be an effective way of authentication. It has been observed that wearable device performances are similar to medical device performances. If the presented system is developed and used for wearable devices such as smartwatches, device authentication security will increase and the impact of daily activities on the authentication system will be minimised.

AUTHOR CONTRIBUTION

Hazal Su Bıçakcı: Conceptualisation; Formal analysis; Investigation; Methodology; Software; Visualisation; Writing – original draft; Writing – review & editing. **Marco Santopietro**: Conceptualisation; Methodology; Software; Visualisation; Writing – original draft; Writing – review & editing. **Richard Guest**: Conceptualisation; Investigation; Methodology; Project administration; Resources, Supervision; Visualisation; Writing – original draft; Writing – review & editing.

ACKNOWLEDGEMENTS

Hazal Su Bıçakcı would like to thank the Ministry of National Education of the Republic of Turkiye for their support and financing.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

Data derived from public domain resources.

ORCID

Hazal Su Bıçakcı https://orcid.org/0000-0002-6514-8634

REFERENCES

1. Shen, T.-W., Tompkins, W.J., Hu, Y.H.: One-lead ecg for identity verification. In: Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][engineering in Medicine and Biology, vol. 1, pp. 62–63. IEEE (2002)
2. Biel, L., et al.: Ecg analysis: a new approach in human identification. IEEE Trans. Instrum. Meas. 50(3), 808–812 (2001). https://doi.org/10.1109/19.930458
3. Israel, S.A., et al.: Ecg to identify individuals. Pattern Recogn. 38(1), 133–142 (2005). https://doi.org/10.1016/j.patcog.2004.05.014
4. Yin, S., et al.: A 1.06-μw smart ecg processor in 65-nm cmos for real-time biometric authentication and personal cardiac monitoring. IEEE J. Solid State Circ. 54(8), 2316–2326 (2019). https://doi.org/10.1109/jssc.2019.2912304
5. Hernando, D., et al.: Validation of heart rate monitor polar rs800 for heart rate variability analysis during exercise. J. Strength Condit Res. 32(3), 716–725 (2018). https://doi.org/10.1519/jsc.0000000000001662
6. Qardiocore: Qardio Smart Wearable ecg/ekg Monitor (2018). [Online]. URL https://www.getqardio.com/qardiocore-wearable-ecg-ekg-%20monitor-
7. Lazaro, J., et al.: Wearable armband device for daily life electrocardiogram monitoring. IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 67(12), 3464–3473 (2020). https://doi.org/10.1109/tbme.2020.2987759
8. Chen, E., et al.: A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation. Heart Rhythm 17(5), 847–853 (2020). https://doi.org/10.1016/j.hrthm.2020.01.034
9. Deanfield, J.E.: Holter monitoring in assessment of angina pectoris. Am. J. Cardiol. 59(7), C18–C22 (1987). https://doi.org/10.1016/0002-9149(87)90191-3
10. Bıçakcı, H.S., et al.: Evaluation of electrocardiogram biometric verification models based on short enrollment time on medical and wearable recorders. In: 2021 International Carnahan Conference on Security Technology (ICCST), pp. 1–6. IEEE, Hertfordshire, U. K (2021)
11. Selvaraj, J., et al.: Classification of emotional states from electrocardiogram signals: a non-linear approach based on Hurst. Biomed. Eng. Online 12(1), 1–18 (2013). https://doi.org/10.1186/1475-925x-12-44
12. Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. IEEE Trans. Pattern Anal. Mach. Intell. 30(12), 2067–2083 (2008). https://doi.org/10.1109/tpami.2008.26
13. Martin, T., Jovanov, E., Raskovic, D.: Issues in wearable computing for medical monitoring applications: a case study of a wearable ecg monitoring device. In: Digest of Papers. Fourth International Symposium on Wearable Computers, pp. 43–49. IEEE (2000)
14. Ingale, M., et al.: Ecg biometric authentication: a comparative analysis. IEEE Access 8, 117853–117866 (2020). https://doi.org/10.1109/access.2020.3004464
15. Hassan, Z., Gilani, S.O., Jamil, M.: Review of fiducial and non-fiducial techniques of feature extraction in ecg based biometric systems. Indian J. Sci. Technol. 9(21), 850–855 (2016). https://doi.org/10.17485/ijst/2016/v9i21/94841
16. Choi, H.-S., Lee, B., Yoon, S.: Biometric authentication using noisy electrocardiograms acquired by mobile sensors. IEEE Access 4, 1266–1273 (2016). https://doi.org/10.1109/access.2016.2548519
17. Krasteva, V., Jekova, I., Abächerli, R.: Biometric verification by cross-correlation analysis of 12-lead ecg patterns: ranking of the most reliable peripheral and chest leads. J. Electrocardiol. 50(6), 847–854 (2017). https://doi.org/10.1016/j.jelectrocard.2017.08.021
18. Paiva, J.S., Dias, D., Cunha, J.P.S.: Beat-id: towards a computationally low-cost single heartbeat biometric identity check system based on electrocardiogram wave morphology. PLoS One 12(7), e0180942 (2017). https://doi.org/10.1371/journal.pone.0180942
19. Sidek, K.A., Khalil, I., Smolen, M.: Ecg biometric recognition in different physiological conditions using robust normalized qrs complexes. In: 2012 Computing in Cardiology, pp. 97–100. IEEE (2012)
20. Kim, H., et al.: Physiology-based augmented deep neural network frameworks for ecg biometrics with short ecg pulses considering varying heart rates. Pattern Recogn. Lett. 156, 1–6 (2022). https://doi.org/10.1016/j.patrec.2022.02.014
21. Lehmann, F., Buschek, D.: Heartbeats in the wild: a field study exploring ecg biometrics in everyday life. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2020)
22. Sriram, J.C., et al.: Activity-aware ecg-based patient authentication for remote health monitoring. In: Proceedings of the 2009 International Conference on Multimodal Interfaces, pp. 297–304 (2009)
23. Mekruksavanich, S., Jitpattanakul, A.: Biometric user identification based on human activity recognition using wearable sensors: an experiment using deep learning models. Electronics 10(3), 308 (2021). https://doi.org/10.3390/electronics10030308
24. Batool, S., Saqib, N.A., Khan, M.A.: Internet of things data analytics for user authentication and activity recognition. In: 2017 Second

International Conference on Fog and Mobile Edge Computing (FMEC), pp. 183–187. IEEE (2017)
25. Li, M., et al.: Multimodal physical activity recognition by fusing temporal and cepstral information. IEEE Trans. Neural Syst. Rehabil. Eng. 18(4), 369–380 (2010). https://doi.org/10.1109/tnsre.2010.2053217
26. Schmidt, P., et al.: Introducing wesad, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction, pp. 400–408 (2018)
27. Lin, J., et al.: An explainable deep fusion network for affect recognition using physiological signals. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2069–2072 (2019)
28. Sarkar, P., Ali, E.: Self-supervised ecg representation learning for emotion recognition. IEEE Trans. Affect. Comput. (2020)
29. Lee, W., Kim, S., Kim, D.: Individual biometric identification using multi-cycle electrocardiographic waveform patterns. Sensors 18(4), 1005 (2018). https://doi.org/10.3390/s18041005
30. Perpiñan, G., et al.: Classification of metabolic syndrome subjects and marathon runners with the k-means algorithm using heart rate variability features. In: 2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA), pp. 1–6. IEEE (2016)
31. Miller, P.E., et al.: Personal identification using periocular skin texture. In: Proceedings of the 2010 ACM Symposium on Applied Computing, pp. 1496–1500 (2010)
32. Abraham, S., Titus, A.: Qrs based arrhythmia classification using k-nn architecture. IJREAT Inter. J. Res. Engin. Adv. Technol. 1(3) (2013)
33. Aggarwal, C.C., Hinneburg, A., Keim, D.A.: On the surprising behavior of distance metrics in high dimensional space. In: International Conference on Database Theory, pp. 420–434. Springer (2001)
34. Vollmer, M., et al.: Simultaneous Physiological Measurements with Five Devices at Different Cognitive and Physical Loads, Version 1.0.0 ed. Physionet (2020)
35. Samson, A.C., et al.: Eliciting positive, negative and mixed emotional states: a film library for affective scientists. Cognit. Emot. 30(5), 827–856 (2016). https://doi.org/10.1080/02699931.2015.1031089
36. Mind Media, B.V.: Mind Media Neuro and Biofeedback Systems (2017). [Online]. URL https://s3.amazonaws.com/cdn.freshdesk.com/data/helpdesk/attachments/production/36102247749/original/UserManual_MM_NeXus-%2010MKII_EN.pdf?response-content-%20type=application
37. Adaam, I., Bittium, B.: (2020). URL https://www.bittium.com/about-bittium/blog/covid-19%20-bittium-faros-echoes-its-name-as-a-cardiac-lighthouse%20-through-mail-ecg-service
38. Bilo, G., et al.: Validation of the somnotouch-nibp noninvasive continuous blood pressure monitor according to the European society of hypertension international protocol revision 2010. Blood Pres. Monit. 20(5), 291–294 (2015). https://doi.org/10.1097/mbp.0000000000000124
39. Hexoskin Health Sensors and AI: Hexoskin Health Sensors & Ai (2013). URL. [Online] https://www.hexoskin.com/pages/health-research
40. Vollmer, M., Bläsing, D., Kaderali, L.: Alignment of multi-sensored data: adjustment of sampling frequencies and time shifts. In: 2019 Computing in Cardiology (CinC), p. 1. IEEE (2019)
41. Goldberger, A.L., et al.: Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. Circulation 101(23), e215–e220 (2000). https://doi.org/10.1161/01.cir.101.23.e215
42. Abdul-Kadir, N.A., Mat Safri, N., Othman, M.A.: Dynamic ecg features for atrial fibrillation recognition. Comput. Methods Progr. Biomed. 136, 143–150 (2016). https://doi.org/10.1016/j.cmpb.2016.08.021
43. Selek, M.B., et al.: The effect of principal component analysis in the diagnosis of congestive heart failure via heart rate variability analysis. Proc. IME H J. Eng. Med. 235(12), 1479–1488 (2021). https://doi.org/10.1177/09544119211036806
44. Ahamed, F., et al.: An intelligent multimodal biometric authentication model for personalised healthcare services. Future Internet 14(8), 222 (2022). https://doi.org/10.3390/fi14080222
45. Yang, J., et al.: Study on ppg Biometric Recognition Based on Multifeature Extraction and Naive Bayes Classifier. Scientific Programming 2021, (2021)
46. Uddin, S., et al.: Comparing different supervised machine learning algorithms for disease prediction. BMC Med. Inf. Decis. Making 19(1), 1–16 (2019). https://doi.org/10.1186/s12911-019-1004-8
47. Brener, V.: 312 Analytics 2022, (2013). URL. [Online] http://www.312analytics.com/decision-trees-vs-neural-%20networks/
48. Gul, A., et al.: Ensemble of a subset of knn classifiers. Adv. Data Anal. Classif. 12(4), 827–840 (2018). https://doi.org/10.1007/s11634-015-0227-5

How to cite this article: Bıçakcı, H.S., Santopietro, M., Guest, R.: Activity-based electrocardiogram biometric verification using wearable devices. IET Biome. 12(1), 38–51 (2023). https://doi.org/10.1049/bme2.12105