Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2019) 27: 3682 – 3698
© TÜBİTAK
doi:10.3906/elk-1901-168

Biometric person authentication framework using polynomial curve fitting-based ECG feature extraction

Şahin IŞIK[1*], Kemal ÖZKAN[1], Semih ERGİN[2]

[1]Department of Computer Engineering, Faculty of Engineering and Architecture, Eskişehir Osmangazi University, Eskişehir, Turkey

[2]Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Eskişehir Osmangazi University, Eskişehir, Turkey

Received: 23.01.2019 • **Accepted/Published Online:** 15.04.2019 • **Final Version:** 18.09.2019

Abstract: The applications of modern biometric techniques for person identification systems rapidly increase for meeting the rising security demands. The distinctive physiological characteristics are more correctly measurable and trustworthy since previous measurements are not appropriately made for physiological properties. While a variety of strategies have been enabled for identification, the electrocardiogram (ECG)-based approaches are popular and reliable techniques in the senses of measurability, singularity, and universal awareness of heartbeat signals. This paper presents a new ECG-based feature extraction method for person identification using a huge amount of ECG recordings. First of all, 1800 heartbeats for each of the 36 subjects have been obtained from the widespread and large MIT-BIH database (MITDB) downloaded from the PhysioBank archive. Then the fiducial points of each heartbeat were determined and fourteen different features were extracted utilizing these fiducial points. Next, the polynomial curve fitting-based dimension reduction technique was employed on the extracted fourteen features. Furthermore, six celebrated classifiers including artificial neural networks (ANNs), decision trees (DTs), Fisher linear discriminant analysis (FLDA), K-nearest neighbors (K-NNs), naive Bayes (NB), and support vector machines (SVMs) were applied for the verification and performance evaluation of the proposed study. Also, as a different classifier, temporal classification and random forest was utilized for a benchmark classification. The highest performance was attained with 95.46% accuracy rate in the case of the SVM classifier. The experimental results emphasize that the proposed ECG-based feature extraction method gives insightful merit for biometric-based person authentication systems.

Key words: Feature extraction, electrocardiogram, curve fitting, biometric identification

1. Introduction

Several biometric authentication techniques have been developed for numerous technological applications such as trustworthy access to restricted areas and the protection of secure documents [1]. The most important issues for these biometric authentication techniques are their recognition performances as well as their easiness in terms of implementations [2]. However, numerous authentication techniques (such as passwords and intelligent ID cards), rather than biometric ones, are used in many different platforms, but they are still not a solution for the problem of stealing and also copying [3]. In addition to this, various physiological signals carry a person's inherent characteristics and they present adequate reliability and long time persistence and usability [4]. One of the physiological signals, electrocardiogram (ECG), is recently preferred as a biometric marker for

*Correspondence: elektrik@tubitak.gov.tr

authentication purposes [1,5,6]. ECG signals are measured by means of one or many electrodes placed on a body surface and they include not only some fundamental physical information such as heart rate but also personally identifiable electrical properties [7]. They are used as a biometric marker since each acquired ECG signal is exclusive for a person [4,8].

There are numerous feature extraction processes successfully achieved on ECG signals and then generally supervised classification is implemented using these features [9,10]. The amplitude- and temporal-based features [2,7,8,11,12] are extracted and used for person identification purposes after an ordinary QRS complex detection process [4,7,13,14]. In addition to these types of features, mel-frequency cepstral coefficient (MFCC) [15], PCA [15], ICA [16], DCT [6], [7], and WT [16] methodologies are performed to extract distinctive features. Moreover, Wilk's lambda method [17], Fourier transform [18], and vectorcardiogram [3] are also methods to derive features from ECG signals. A combinational feature extraction scheme is implemented utilizing both cepstral and temporal information [19]. Hammad et al. applied two cancelable techniques for implementing a person authentication process using ECG signals [20]. The first one is an improved bio-hashing while the other one is a matrix operation technique. Louis et al. proposed one-dimensional multiresolution local binary patterns (1DMRLBP), which was carried out as an online feature extraction method for measured ECG signals. In their methodology, all decision thresholds can be adaptively updated [21]. Hejazi et al. performed nonfiducial methodologies based on an autocorrelation (AC) together with linear dimension reduction to derive noteworthy ECG features [22]. Zhang et al. proposed a fusion of fiducial- and nonfiducial-based ECG features to obtain more discriminative ECG features [23]. Therefore, they increased the stability for an authentication process. In addition, Ergin et al. [24] used the ensemble of features (QRS features, time-domain features, wavelet transform features, and power spectral density features) to classify 18 healthy people using only 2-s ECG recordings, and Gurkan et al. [25] considered the combination of AC/DCT features, MFCC features, and QRS beat information. Also, an extended Kalman filtering (EKF) framework was proposed to extract ECG features on fiducial points [26].

One of the most popular methods for ECG based biometric recognition is to utilize the distance measure [2,4,10,12,14,27] in the literature. The nearest neighbor classifier [1,15,28,29,30,31], neural networks [6], and decision-based neural networks [11,12] are successfully used for person identification. Support vector machine [3,6,19], template matching with the correlation coefficient [4,5], radial basis function [13,16], Gaussian mixture model (GMM) [19], and similarity or dissimilarity measure [9] are the additional recently proposed methods. Linear discriminant analysis [7,8], wavelet-based classification with thresholding [5,32], rank classifier [17], and clustering [33] are the other classifiers performed for person identification. Tantawi et al. suggested discrete biorthogonal wavelets in which the RR intervals of ECG waveforms are decomposed. These wavelet features were then used with a radial basis function (RBF) neural network for classification [34]. Tan and Perkowski announced a two-stage classifier combining random forest and wavelet distance measure through a probabilistic threshold criterion. They applied their method on a biosensor-integrated mobile device [35]. Coutinho et al. proposed the Ziv–Merhav cross-parsing algorithm with the estimator of complexity [36]. Karimian et al. derived a novel key generation methodology that produces keys from real-valued ECG features with high reliability and entropy [37]. Additionally, a comparative analysis for ECG biometric authentication performances of several studies was given in [38].

One can easily observe that the majority of the above-mentioned methodologies have been conducted on low-volume ECG data. Since the problem of high-volume ECG recordings was not examined, a novel and innovative technique is vitally needed to be implemented on big datasets with a high number of heartbeats. Taking this fact into consideration, a new ECG-based feature extraction method is proposed for person identification in this study using a huge amount of ECG recordings. First of all, 1800 heartbeats for each of the 36 subjects were obtained. Then the fiducial points of each beat were determined and 14 different features were computed utilizing from these fiducial points. Next, the curve fitting-based dimension reduction technique was employed on the extracted 14 features. For the verification and performance evaluation of the proposed study, seven famous classifiers including artificial neural networks (ANNs), decision trees (DTs), Fisher linear discriminant analysis (FLDA), K-nearest neighbors (kNNs), naive Bayes (NB), random forest (RF), and support vector machines (SVMs) were applied. Also, a time-series based classifier, called temporal classifier (TEMP) [39], has been utilized in order to constitute a comparative classification. The highest performance of our proposed study was attained as 95.46% accuracy rate in the case of the SVM classifier.

The rest of this paper is organized such that the definitions of the extracted 14 features and the concept of polynomial curve fitting are given in Sections 2 and 3, respectively. The fourth section includes concise descriptions of each classifier while the fifth section presents the information about the dataset, proposed feature extraction procedure, and experimental results. The last section summarizes conclusions revealed from the proposed study.

2. Methods and tools

Figure 1 shows the proposed ECG-based authentication system with essential steps. The main motivation behind this approach is determining the combination of discriminative features and a robust classifier in order to effectively authenticate individuals. After the data acquisition stage, the ECG signal is segmented into different heartbeat parts. The second step is the preprocessing phase, which is challenging and important to enhance the performance of identification. The preprocessing stage consists of two phases; the first is removing noise factors from ECG data by a smoothing operation and then applying a differentiation procedure to shed light on the location of fiducial points. Once the location of the fiducial points is spotted, the raw ECG features are determined to encode the electrical activity of heartbeats into meaningful statistical values. Then the curve fitting procedure is applied to the ECG features to improve the discriminative capability of the system. In the last phase, a generalizable model is investigated by conducting training and test simulations on extracted curve fitting coefficients.

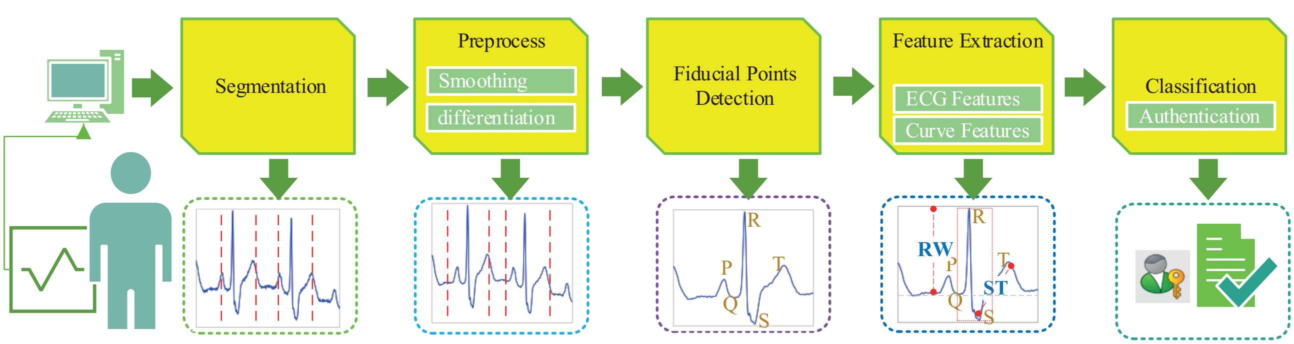

Figure 1. Proposed ECG-based authentication system.

2.1. ECG-based features

The main contribution of ECG features in person identification is to provide a meaningful interpretation of the electrical activity of a heart. The invention of the first practical electrocardiogram dates back to the study of Einthoven in 1903. Also, the first experiment on person identification with numerical ECG features was made by Biel et al. [40]. In that study, common elements of a normal ECG beat, called fiducial points, were specified by considering that a heartbeat occurs by states of depolarization and repolarization of the muscle fibers.

Useful information about beat rhythms can be obtained by recording the electrical impulses reflected from the pumping activity of a heart. With this aspect, any ECG signal can be constituted from the depolarization and repolarization of muscle fibers. While depolarization indicates the P-wave (atrial depolarization) and QRS-wave (ventricles depolarization), repolarization refers to the T-wave and U-wave (ventricular repolarization) [40]. Usually, the duration, amplitude, and trigonometric interpretations of waves in an ECG signal have been utilized as the discriminative features of heartbeats. In this paper, the ability of features that are capable of categorizing the heart characteristics is first considered. First of all, the fiducial points of a processed beat are

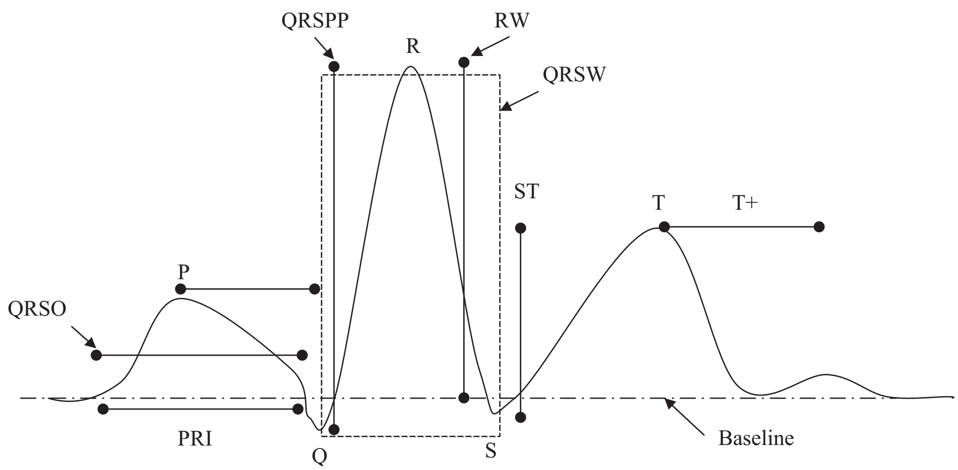

Figure 2. The overview of some fiducial points.

detected with a new implementation of the Pan–Tompkins algorithm [41] for QRS detection in ECG complexes. The value of a fiducial point is explored after performing a simple differentiation procedure on a smoothed ECG signal. Once the QRS complex is obtained, the remaining fiducial points (P and T) are attained in order to determine to borderline of the ECG beat. Then slope, duration, and amplitudes of consecutive fiducial points are computed so that the features given in Table 1 are extracted. An ECG signal is illustrated for better understanding and interpreting the ECG features in Figure 2. Moreover, brief descriptions of the extracted ECG features are given in Table 1. One may infer from Table 1 that totally 14 features have been calculated for the biometric representation of a heartbeat of a person. Although the performances of PRI, RWD, P+, QRSPP, RW, ST, T+, QRSW, QRSO, and STS features have been evaluated in some previous studies [40,42], the evaluation of the remaining four features is investigated here for the first time in the literature. In this study, a biometric identification process is operated based on all of the computed 14 features.

Table 1. The utilized ECG features.

Detail	Feature	Description
Features in the literature [24, 40]	PRI	Duration between the beginning of P wave and QRS complex
	RWD	Elapsed time between the beginning and end of R wave
	P+	Difference between P point and other subsequent point where ECG signal rises again
	QRSPP	Difference between R and Q points in QRS complex in terms of amplitude
	RW	Height of R wave from the baseline
	ST	Difference between S and T points in terms of amplitude
	T+	Difference between T point and the subsequent point where ECG signal rises again
	QRSW	Area of the rectangle drawn on QRS complex using Q, R, and S points
	QRSO	Duration between the beginning and end of P wave
	STS	Angle of the line drawn from S point to T point
Proposed new features	PTS	Angle of the line drawn from P point to T point
	PRS	Angle of the line drawn from P point to R point
	QTS	Angle of the line drawn from Q point to T point
	QSS	Angle of the line drawn from Q point to S point

2.2. Polynomial curve-fitting

Technically, curve fitting provides a robust linear or nonlinear model that is able to characterize the relationship between independent and dependent parameters. To avoid underfitting and overfitting cases, one may use a polynomial or linear model based on the distribution of data. Since there is a weak connection between independent parameters, called heartbeat features, we applied a nonlinear model driven by the concept of the polynomial curve fitting procedure. The motivation behind this study is that we explore the effectiveness of using coefficients of a fitted polynomial model with the purpose of person identification. The underline idea of the utilized nonlinear model is described with the following equations.

Given two variables ($x_1\, and\, x_2$), a second-order polynomial model can be represented with coefficients for the developed nonlinear model.

$$y = a_0 + a_1 x + a_2 x^2 + e \tag{1}$$

The aim is to minimize the residual value, namely the error (e) between predicted and expected values [42]:

$$e_{sum} = \sum_{i=1}^{m} \left(y_i - a_0 - a_1 x_1 - a_2 {x_2}^2 \right)^2 \tag{2}$$

The derivation of model coefficients relied on taking the derivative of both sides of the equation for each coefficient in Eq. (3):

$$\begin{aligned}\frac{\partial e_{sum}}{\partial a_0} &= -2\sum_{i=1}^{m}\left(y_i - a_0 - a_1 x_{\mathrm{i}} - a_2 x_{\mathrm{i}}^2\right)\\ \frac{\partial e_{sum}}{\partial a_1} &= -2\sum_{i=1}^{m} x_i\left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)\\ \frac{\partial e_{sum}}{\partial a_2} &= -2\sum_{i=1}^{m} x_i^2\left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)\end{aligned} \tag{3}$$

Eventually the model coefficients are determined after replacing the parameters as shown in Eq. (4).

$$\begin{bmatrix} n & \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 \\ \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 & \sum_{i=1}^{m} x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \\ \sum_{i=1}^{m} x_i^2 y_i \end{bmatrix} \tag{4}$$

By considering the motivation from Eq. (4), the mth-order polynomial fitting methodology can be realized as demonstrated with Eq. (5). As one can infer from Eq. (5), it is required to solve the system of linear equations including (m+1) equations with (m+1) unknowns in order to approximate the data with an mth-order polynomial model. In this study, the coefficients of 2nd, 3rd, 4th, 5th, and 6th order polynomial curve fitting model are evaluated on the basis of biometric authentication accuracy and the computational cost of algorithm.

$$\begin{bmatrix} n & \sum_{i=1}^{m} x_i & \ldots & \sum_{i=1}^{m} x_i^2 \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \ldots & \sum_{i=1}^{m} x_i^3 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \sum_{i=1}^{m} x_i^m & \sum_{i=1}^{m} x_i^{m+1} & \ldots & \sum_{i=1}^{m} x_i^{m+m} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ . \\ . \\ . \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \\ . \\ . \\ . \\ \sum_{i=1}^{m} x_i^m y_i \end{bmatrix} \tag{5}$$

2.3. Feature extraction

Several previous studies were focused on feature extraction strategies using the projection methods, which are PCA [40], LDA [32], ICA [43], and SVD [44]. Also, other extraction strategies were based on the impacts of feature transformation methods [45,46] and statistical techniques [47]. With a different point of view, the potential impact of polynomial curve fitting methodology is investigated to obtain discriminative features in this study. Detailed information about the proposed feature extraction phase is presented in Figure 3. In the feature extraction process, first of all, a smoothing procedure is performed to capture the noise trend within ECG signals. Then the identified noise is removed by the Savitzky-–Golay filter [48] to effectively eliminate the noise parts in an ECG signal. The polynomial order of our Savitzky–Golay filter is 6 and this filter is directly

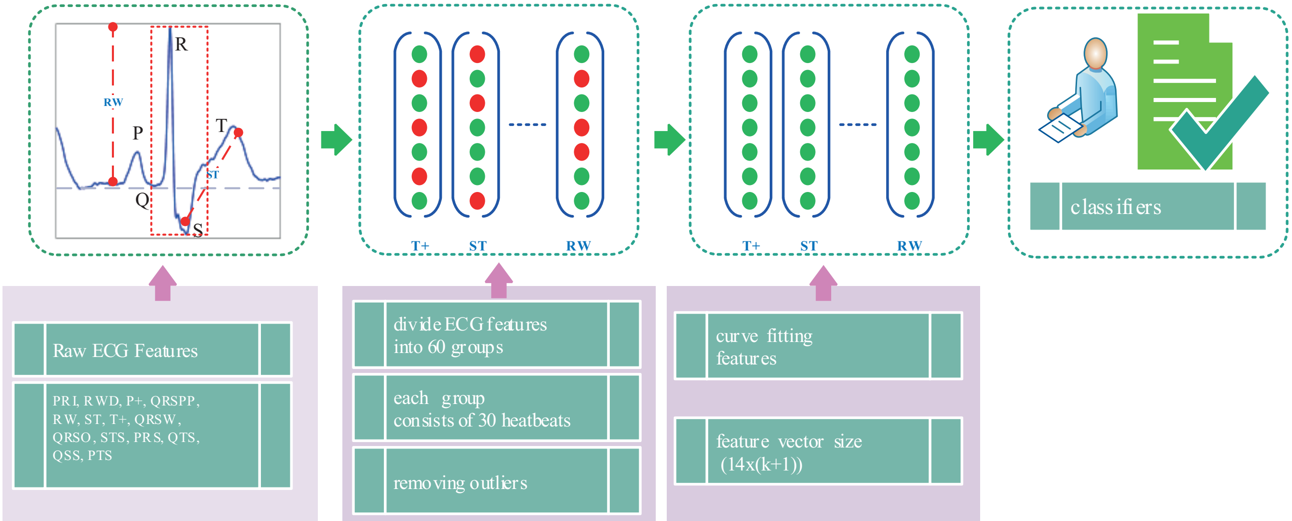

Figure 3. Curve fitting-based feature extraction.

applied to the ECG samples without any framing methodology. Once the noise is removed from ECG data, a peak detection using a simple differentiation process is introduced to find the P, QRS complex, and T peak. After the detection of these five fiducial points (P, Q, R, S, and T points), a total of 1800 heartbeats are found in the ECG signal of each subject. One should note that the proposed algorithm detects the P, Q, R, S, and T values after applying a simple differentiation procedure, which locates the maximum and minimum values of the heartbeat signal. Next, all of these heartbeats are grouped so that each subject has 60 groups and each group consists of 30 heartbeats. The 14 features described in Table 1 are acquired from each heartbeat in each group. Therefore, 30 values are obtained for each feature type in each group.

When the obtained 30 values are examined, it is easily seen that a preprocessing step is vital to remove the outliers that ruin the general trend of data. The reason for outliers may include a physiological factor such as ectopics, but they should be ignored in order to find a specific curve for a particular subject. Therefore, the values that are less than 1.5 times the average of 30 values are removed after the normalization of values into the range of [0-1] using the zero-score normalization technique [44]. Moreover, in order to reduce the number of values for each feature type, the polynomial curve fitting strategy is applied to the obtained 30 values. The kth-order polynomial curve is fitted on the 30 obtained values so that each group is represented with (k+1) coefficients for each feature type. The exponent k is changing from 2 to 6. Thus, each 60 groups for each subject are represented with (14x(k+1))-dimensional feature vectors. For example, if a 2nd degree polynomial curve is fitted, the dimension of each feature vector of any subject will be 42 (=14x3). The same feature extraction procedure is employed for all subjects so that every subject has 60 feature vectors whose dimensions are (14x(k+1)).

The demonstration of general steps of feature extraction with polynomial curve fitting is given in Figure 4. Due to space limitations, only the QRSO features obtained from 30 heartbeats are plotted. In Figure 4a, an outlier appears between the –250 and –300 amplitude values, whereas the general trend in the data varies between 0 and –50 amplitude values. It is widely known that outliers ruin the general trend of data. Since outliers increase the sum of the error square measure and lead to incorrect statistical results, it is definitely

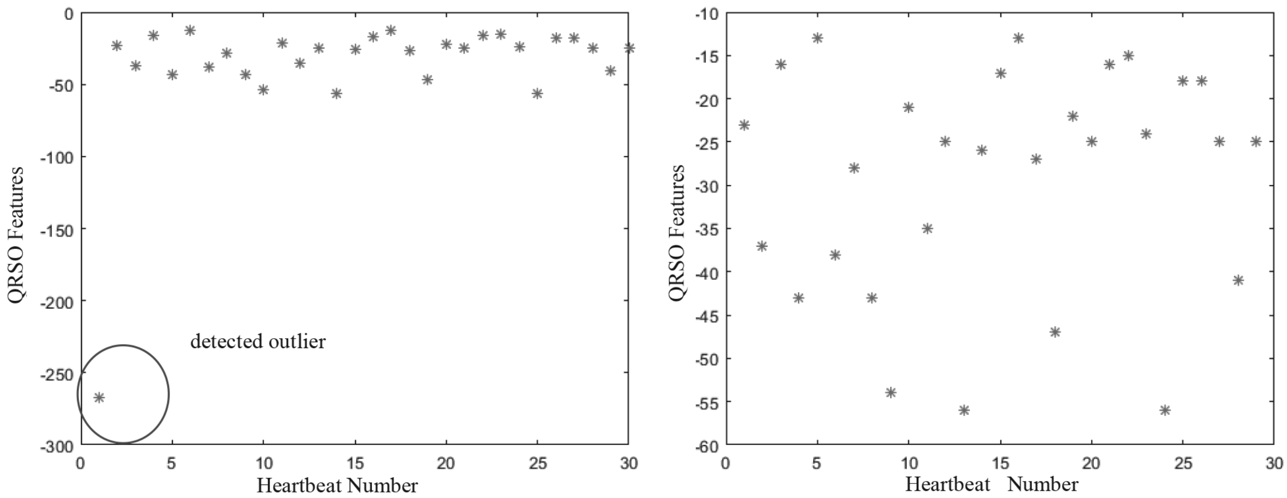

Figure 4. Thirty values in a group for the QRSO feature: (a) before the outlier rejection; (b) detected outliers are removed and nonoutlier features are zoomed (a 5th degree curve is fitted on nonoutlier features).

needed to discard them by using a predefined smart rule or a statistical algorithm. Therefore, we have utilized the z-score algorithm [49] to eliminate outliers. Later, the 5th degree polynomial curve is fitted on the remaining nonoutlier feature values as illustrated in Figure 4b. For clarification purposes, the nonoutlier feature values in Figure 4a are zoomed and sketched in Figure 4b. In this way, it is believed that the computational load will be reduced by excluding redundant values with an effortless process.

Eventually, 42-, 56-, 70-, 84-, and 98-dimensional feature vectors are constructed by fitting the 2nd, 3rd, 4th, 5th, and 6th degree polynomial curves, respectively. The obtained coefficients for the 2nd, 3rd, 4th, 5th, and 6th degree curves are presented in Table 2 for a 30-heartbeat group of the QRSO feature.

Table 2. The average of coefficients for five different fitted curves for the QRSO feature.

Degree	Coefficients
2	0.0082, -0.0650, -30.6535
3	-0.0068, 0.3137, -3.7931, -20.5527
4	0.0002, -0.0175, 0.5228, -5.2441, -18.0462
5	0.0001, -0.0052, 0.1260, -1.1352, 2.3680, -27.4652
6	-0.0000, 0.0005, -0.0197, 0.3641, -3.0055, 8.5937, -33.4288

2.4. Classifiers

In pattern recognition systems, choosing the right classifier to obtain successful results depends on some parameters including characteristic of the problem, computational cost, dimension, and type of data. Specifically, if one would like to classify binary features, the binary naive Bayes can be performed on the feature set. In the proposed study, we have utilized well-known classifiers for the following reasons: the features derived from an ECG signal are statistical, and these classifiers have been preferred in most of ECG systems for anomaly and person identification tasks. In this study, we have analyzed the capability of popular classifiers including artificial neural networks (ANNs), decision trees (DTs), Fisher linear discriminant analysis (FLDA), K-nearest neighbors (kNNs), naive Bayes (NB), random forest (RF), and support vector machines (SVMs).

In the case of conducting training simulations on the SVM classifier, the strategy of "one to all" procedure have considered for multi-class problems. Due to the nonlinearity of data, we have applied kernel discriminant analysis (KDA) to transform the ECG data into linear form. Then the KDA-projected features are classified with the SVM classifier by using the polynomial kernel. The order of the polynomial kernel function is 2 and the Gaussian parameter is 1. For the K-NN classifier, the distance metric is standardized Euclidean and the number of nearest neighbors is specified as 1. When training the ANN classifier, the scaled conjugate gradient method is utilized to update weights of the network through backpropagation. The transfer function in layers is arranged as softmax.

3. Experimental study

3.1. Dataset

The popular and large MIT-BIH Arrhythmia (MITDB) database is downloaded from the PhysioBank website [50] and used to emphasize the performance of the proposed method. MITDB contains 48 records belonging to 47 subjects, 25 males and 22 female ranging from 23 years to 89 years old. There are two records per subject, and the duration of each record is about 30 min with sampling frequency of 360 Hz. The ECG signals of 36 subjects are used in the experiments. The reason for the selection of 36 subjects rather than using 47 subjects is that the duration of ECG signals for the remaining 11 subjects is less than 30 min. Therefore, those subjects are removed in order to obtain equal prior probability for each subject class before training. In this way, any domination of a class over a different class can be prevented for fair comparison.

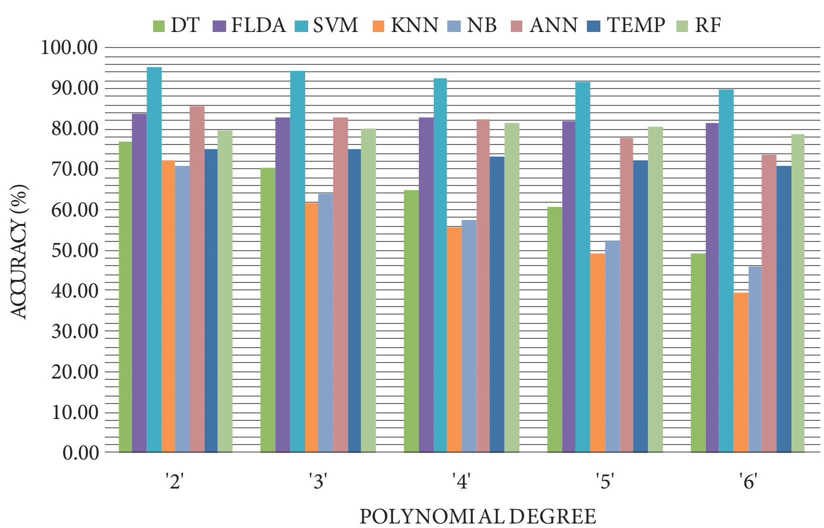

Figure 5. The average recognition accuracies (%) attained by each classifier employing four-fold cross-validation.

3.2. Classification results

The performance of the proposed feature extraction method on ECG signals for subject authentication is evaluated using eight different celebrated classifiers: ANN, DT, FLDA, K-NN, NB, TEMP, RF, and SVM. Recall that the main objective of the proposed method is to analyze the impact of a curve fitting-based feature extraction strategy on the identification of subjects using ECG data signals. The use of the polynomial curve fitting methodology on ECG-based features is considered as an important contribution to the ECG literature. Four-fold cross-validation is operated on the feature vectors in order to objectively quantify the performance evaluation of the proposed study. Specifically, 45 feature vectors of each subject are chosen for training and the remaining 15 feature vectors of the corresponding subject are used in the test stage for each cross-validation step. Figure 5 summarizes the performance attained from each classifier. The recognition accuracies given in Figure 5 are the average accuracies of all cross-validation steps. Upon inspecting the obtained results, one can easily note that the SVM classifier presents the most satisfactory accuracy rates in comparison with other classifiers. Obviously, it can be clearly observed that the best performance is obtained with 2nd degree polynomial curve coefficients. The highest accuracy rates for the DT, FLDA, SVM, KNN, NB, ANN, TEMP, and RF classifiers are 76.71%, 83.98%, 95.46%, 72.04%, 70.74%, 85.60%, 75.09%, and 81.34%, respectively.

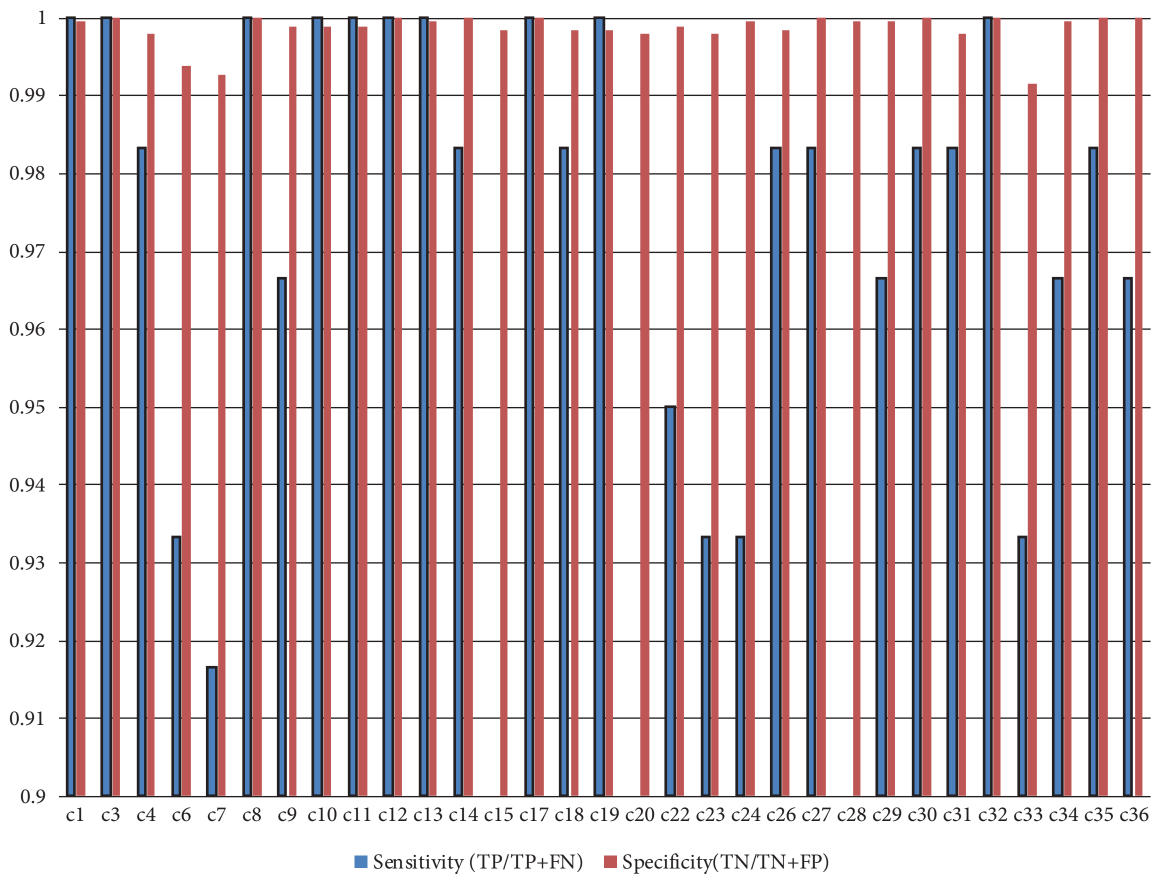

Figure 6. Sensitivity and specificity of the proposed authentication model.

Additionally, the correctly and incorrectly classified results are demonstrated to objectively verify the performance of the proposed study. The obtained statistical results are demonstrated in Figure 6, which shows the specificity and sensitivity results derived from the confusion matrix for the coefficients of the 2nd degree polynomial curve. Only the best sensitivity scores that are greater than 90% are plotted in Figure 6. For this purpose, the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) rates corresponding to each class is computed by considering the confusion matrix for the multiclass case. In this study, the "specificity" metric refers to the proportion of TN to the sum of TN and FP ($\mathbf{TN/(TN + FP)}$) and "sensitivity" indicates the proportion of TP to the sum of TP and FN ($\mathbf{TP/(TP + FN)}$). When one considers Figure 6, it is obvious that subjects C1, C3, C8, C10, C11, C12, C13, C17, C19, and C32 are perfectly identified with 100% recognition performance in terms of sensitivity. Therefore, the highest value of the sensitivity score is measured as 1.

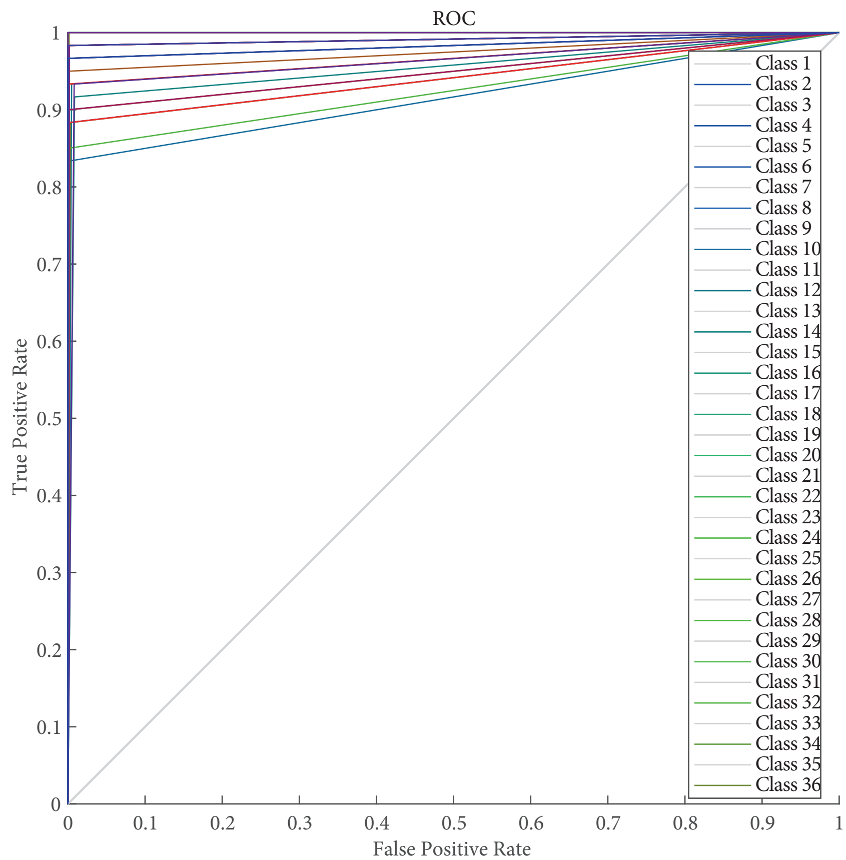

Figure 7. AUC results of proposed model.

Moreover, the area under the curve (AUC) results are determined by constructing receiver operating characteristic (ROC) curves. Figure 7 summarizes the performance of the proposed method in terms of AUC values. A higher AUC value indicates better performance. Upon inspecting the results, we can observe that nice AUC results are acquired for the SVM classifier in the case of 2nd degree polynomial curve coefficients.

To investigate the capability of the utilized curve fitting based features, we compare the performance of the proposed method with other approaches. Table 3 presents the performance of each study in terms of accuracy and F-score metrics. For a fairer benchmark, one can compare the results of the proposed method with a comprehensive study [51] since the same database (MITDB) and classifier (SVM) are utilized. In the referred study, the authors conducted an experimental study of 195 subjects with 128 heartbeats for each subject. They observed accuracy rates of 95% and 34% when performing the two feature extraction methods, which were proposed by Yu and Chen [46] and Chou (2008) [43], respectively. Again, the studies in [40], [52], [53], and [54] considered their own ECG datasets when it comes to evaluating the capability of ECG features by means of biometric-based person identification. Focusing on their performances, one may note that the discriminative potential of the phase space trajectory of an ECG signal is higher than that of the other feature extraction

Table 3. The comparison of the proposed work to the literature.

Reference	Database	# of classes	Type of features	Classifier	Results (criteria)
[24]	The MIT-BIH Normal Sinus Rhythm Database	18	QRS	C4.5 Decision Tree	0.96 (F-score)
			TD		0.88 (F-score)
			WT		0.63 (F-score)
			PSD		0.84 (F-score)
			QRS, TD, WT, PSD		0.97 (F-score)
			QRS	BayesNet	0.94 (F-score)
			TD		0.80 (F-score)
			WT		0.54 (F-score)
			PSD		0.79 (F-score)
			QRS, TD, WT, PSD		0.96 (F-score)
[51]	Composition of MITDB, MITSUP, NSRDB, and European ST-T (EDB)	193	WT, statistical techniques, RR interval, ICA, RR-predecessor	SVM	More than 95% (accuracy), not exactly mentioned in the study
[40]	Own database: 12-lead rest ECG recordings	22	12 fiducial points	Generative model classifier (GMC)	100 (Accuracy)
[52]	Two own databases: data from males and females between the ages of 22 and 48 and FDA-approved ECG device	104	15 fiducial points	LDA	91.00 (Accuracy)
[53]	Own database	45	WT	Euclidean distance measure	95.71 (Accuracy)
[54]	Own database: composed from the anterior lead, lateral lead, posterior lead	100	Phase space trajectory of an ECG signal	Mutual nearest point distance (MNPD)	99.00 (Accuracy)
				Normalized spatial correlation (nSC)	98.00 (Accuracy)
				Mutual nearest point match (MNPM)	98.00 (Accuracy)
Proposed work	MIT-BIH Arrhythmia Database	36	Proposed polynomial curve fitting-based ECG features	DT	76.71 (Accuracy)
				FLDA	83.98 (Accuracy)
				SVM	95.46 (Accuracy)
				KNN	72.04 (Accuracy)
				NB	70.74 (Accuracy)
				ANN	85.60 (Accuracy)
				TEMP	75.09 (Accuracy)
				RF	81.34 (Accuracy)

methodology. Although the performance of [40] accounts for a 100% accuracy rate, the size of utilized dataset is relatively small since it is composed of 12-lead rest ECG recordings. One can emphasize that the proposed framework provides valuable identification scores that are better results than some other methods. Among the utilized classifiers, the performances of SVM and ANN are superior with 95.46% and 85.60% accuracy scores achieved in the case of distinguishing individuals from their ECG recordings.

4. Discussion

One of the most important aspects for daily life is living in a secure environment. Therefore, it is considerably taken into account for many biometric applications. One can say that the other biological authentication signatures such as speech, faces, fingerprints, etc. are utilized in different areas; ECG-based authentication can be a respectable alternative for identification of a subject in a security system. With this purpose, we have investigated the merits of choosing the ECG for authentication processes by ignoring some subjects related to the proper recording of the ECG signal after body contact. When the recognition results are examined, the most successful classifiers are SVM and ANN. This is not surprising since they are accepted as the most efficacious classifiers in the pattern recognition domain. The question arising here is why the other classifiers such as FLDA and DT give lower accuracy rates compared to SVM and ANN. The reason for their failure is that the ECG signals of healthy subjects are similar to each other, so more discriminative between-class scatter cannot be evaluated. Since the eigenvectors corresponding to the highest eigenvalues of between-class scatter determine the directions where data distributions have the largest possible variances, i.e. as much variability in the data as possible, the subject classes cannot be distinguished from each other. The DT classifier naturally finds the correct node values by itself. However, for the same reason, the similarity between ECG signals of the healthy subjects affects the inherent nature of discriminative node values in a decision tree. If the degrees for fitted curve polynomials are benchmarked, the highest recognition rates are evaluated for second degree polynomials. This consequence has a much simpler explanation than that of classifier analysis, and it is directly related to the data trend for one subject after the outlier rejection process. The outlier rejection step is vital in this study, and if it is successfully implemented, the data trend is not ruined; that is, the maximum and minimum values for an ECG parameter are very close to each other. Therefore, a fitted curve polynomial has fewer local maxima and minima and so it has fewer roots for the first derivation. This outcome clearly implies that lower degrees for fitted curve polynomials are needed by means of accurate verification of the person.

5. Conclusion

In this study, a new ECG-based feature extraction method for person identification is proposed. The fiducial points of heartbeats for each person are specified and fourteen features are computed. Then the polynomial curve fitting-based dimension reduction technique is employed and finally the seven popular classifiers are applied with a four-fold cross-validation for the performance analysis of the proposed study. In the dimension reduction stage of feature vectors, the polynomial curve functions with five different degrees are fitted to evaluate the system performance. Upon inspecting the curve coefficients in the sense of their performances for each classifier, one can easily conclude that the coefficients generated for 2nd degree polynomial curves are more discriminative and representative than the other fitted polynomial curves. The reason for this is that the number of nearly zero-valued coefficients is computed when using the 3rd, 4th, 5th, and 6th degree polynomial curves. The feature vectors including these approximately zero-valued coefficients do not have satisfactory discriminative power and disclose the interclass differences. As a natural outcome, for the above-mentioned reason, the performances of the classifiers decrease when the number of coefficients for fitted polynomial curves is increased as explicitly shown in Figure 5. When a benchmark is applied to the classifier performances, the SVM with KDA kernel function surpasses the other classifiers. This consequence is not surprising since the SVM classifier is one of the most successful classifiers for many pattern recognition problems if the order of the kernel function and the value of Gaussian parameter are successfully selected [55]. If a comparison is made between the given results and ours, the performance of the proposed method, with a maximum of 95.46% accuracy rate, is superior to some given techniques. Also, the experimental results prove that the proposed ECG-based feature extraction method using polynomial curve fitting overcomes the problem of long ECG recording duration. The implemented biometric person authentication framework not only promises an effortless and successful identification system but also has insightful merit. As a future work, the performance of the developed feature extraction methodology will be improved by using different modeling approaches.

References

[1] Chen CK, Lin CL, Chiu YM. Individual identification based on chaotic electrocardiogram signals. In: IEEE Conference on Industrial Electronics and Applications; Beijing, China; 2011. pp. 1771-1776.

[2] Gahi Y, Lamrani M, Zoglat A, Guennoun M, Kapralos B et al. Biometric identification system based on electrocardiogram data. In: New Technologies, Mobility and Security; Tangier, Morocco; 2008. pp. 1-5.

[3] Lee J, Chee Y, Kim I. Personal identification based on vectorcardiogram derived from limb leads electrocardiogram. Journal of Applied Mathematics 2012; 2012: 904905. doi: 10.1155/2012/904905.

[4] Sasikala P, Wahidabanu RSD. Identification of individuals using electrocardiogram. International Journal of Computer Science and Network Security 2010; 10 (12): 147-153. doi: 10.1186/s12938-015-0072-y.

[5] Chan AD, Hamdy MM, Badre A, Badee V. Wavelet distance measure for person identification using electrocardiograms. IEEE Transactions on Instrumentation and Measurement 2008; 57 (2): 248-253. doi: 10.1109/TIM.2007.909996.

[6] Tawfik MM, Kamal HST. Human identification using QT signal and QRS complex of the ECG. Online Journal of Electrical and Electronic Engineering 2011; 3 (1): 383–387.

[7] Wang Y, Agrafioti F, Hatzinakos D, Plataniotis KN. Analysis of human electrocardiogram for biometric recognition. EURASIP Journal on Advances in Signal Processing 2008; 2008 (1): 148658. doi: 10.1155/2008/148658.

[8] Israel SA, Irvine JM, Cheng A, Wiederhold MD, Wiederhold BK. ECG to identify individuals. Pattern Recognition 2005; 38 (1): 133-142. doi: 10.1016/j.patcog.2004.05.014

[9] Fang SC, Chan HL. Human identification by quantifying similarity and dissimilarity in electrocardiogram phase space. Pattern Recognition 2009; 42 (9): 1824-1831. doi: 10.1016/j.patcog.2008.11.020

[10] Wübbeler G, Stavridis M, Kreiseler D, Bousseljot RD, Elster C. Verification of humans using the electrocardiogram. Pattern Recognition Letters 2007; 28 (10): 1172-1175. doi: 10.1016/j.patrec.2007.01.014

[11] Shen TW, Tompkins WJ, Hu YH. 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; Houston, TX, USA; 2002. pp. 62-63. doi: 10.1109/IEMBS.2002.1134388

[12] Shen TW. Biometric identity verification based on Electrocardiogram (ECG). PhD, University of Wisconsin, Madison, WI, USA, 2005.

[13] Sidek KA, Khalil I. Automobile driver recognition under different physiological conditions using the electrocardiogram. In: IEEE Conference on Computing in Cardiology; Hangzhou, China; 2011. pp. 753-756.

[14] Kim KS, Yoon TH, Lee JW, Kim DJ, Koo HS. A robust human identification by normalized time-domain features of electrocardiogram. In: IEEE Engineering in Medicine and Biology 27th Annual Conference; Shanghai, China; 2005. pp. 1114-1117. doi: 10.1109/IEMBS.2005.1616615.

[15] Sufi F, Khalil I, Hu J. ECG-based authentication. In: Stavroulakis, P, Stamp M (editors). Handbook of Information and Communication Security. Berlin, Germany: Springer, 2010. pp. 309-331.

[16] Ye C, Coimbra MT, Kumar BV. Investigation of human identification using two-lead electrocardiogram (ECG) signals. In: Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems; Washington, DC, USA; 2010. pp. 1-8. doi: 10.1109/BTAS.2010.5634478

[17] Chan AD, Hamdy MM, Badre A, Badee V. Person identification using electrocardiograms. In: Canadian Conference on Electrical and Computer Engineering; Ottawa, Canada; 2006. pp. 1-4. doi: 10.1109/CCECE.2006.277291

[18] Saechia S, Koseeyaporn J, Wardkein P. Human identification system based ECG signal. In: TENCON 2005 - 2005 IEEE Region 10 Conference; Melbourne, Australia; 2005. pp. 1-4. doi: 10.1109/TENCON.2005.300986

[19] Li M, Narayanan S. Robust ECG biometrics by fusing temporal and cepstral information. In: 20th International Conference on Pattern Recognition; İstanbul, Turkey; 2010. pp. 1326-1329. doi: 10.1109/ICPR.2010.330

[20] Hammad M, Luo G, Wang K. Cancelable biometric authentication system based on ECG. Multimedia Tools and Applications 2019; 78 (2): 1857-1887. doi: 10.1007/s11042-018-6300-2

[21] Louis W, Komeili M, Hatzinakos D. Continuous authentication using one-dimensional multi-resolution local binary patterns (1DMRLBP) in ECG biometrics. IEEE Transactions on Information Forensics and Security 2016; 11 (12): 2818-2832. doi: 10.1109/TIFS.2016.2599270

[22] Hejazi M, Al-Haddad SAR, Singh YP, Hashim SJ, Aziz AFA. ECG biometric authentication based on non-fiducial approach using kernel methods. Digital Signal Processing 2016; 52: 72-86. doi: 10.1016/j.dsp.2016.02.008.

[23] Zhang Y, Gravina R, Lu H, Villari M, Fortino G. PEA: Parallel electrocardiogram-based authentication for smart healthcare systems. Journal of Network and Computer Applications 2018; 117: 10-16. doi: 10.1016/j.jnca.2018.05.007.

[24] Ergin S, Uysal AK, Gunal ES, Gunal S, Gulmezoglu MB. ECG based biometric authentication using ensemble of features. In: IEEE 9th Iberian Conference on Information Systems and Technologies; Barcelona, Spain; 2014. pp. 1-6. doi: 10.1109/CISTI.2014.6877089

[25] Gurkan H, Guz U, Yarman BS. A novel human identification system based on electrocardiogram features. In: IEEE International Symposium on Signals, Circuits and Systems; Iasi, Romania; 2013. pp. 1-4. doi: 10.1109/ISSCS.2013.6651266

[26] Ting CM, Salleh SH. ECG based personal identification using extended Kalman filter. In: IEEE 10th International Conference on Information Science, Signal Processing and Their Applications; Kuala Lumpur, Malaysia; 2010. pp. 774-777. doi: 10.1109/ISSPA.2010.5605516

[27] Kaul A, Arora AS, Chauhan S. ECG based human authentication using synthetic ECG template. In IEEE International Conference on Signal Processing, Computing and Control; Waknaghat Solan, India; 2012. pp. 1-4.

[28] Wang J, She M, Nahavandi S, Kouzani A. Human identification from ECG signals via sparse representation of local segments. IEEE Signal Processing Letters 2013; 20 (10): 937-940. doi: 10.1109/LSP.2013.2267593

[29] Silva H, Lourenço A, Canento F, Fred AL, Raposo N. ECG biometrics: principles and applications. In: Proceedings of the International Conference on Bio-inspired Systems and Signal Processing - Biosignals - INSTICC; Spain; 2013. pp. 215-220.

[30] Kouchaki S, Dehghani A, Omranian S, Boostani R. ECG-based personal identification using empirical mode decomposition and Hilbert transform. In: IEEE 16th CSI International Symposium on Artificial Intelligence and Signal Processing; Shiraz, Iran; 2012. pp. 569-573. doi: 10.1109/AISP.2012.6313811

[31] Venkatesh N, Jayaraman S. Human electrocardiogram for biometrics using DTW and FLDA. In: 20th IEEE International Conference on Pattern Recognition; İstanbul, Turkey; 2010. pp. 3838-3841. doi: 10.1109/ICPR.2010.935

[32] Fatemian SZ, Hatzinakos D. A new ECG feature extractor for biometric recognition. In: IEEE 16th International Conference on Digital Signal Processing; Santorini, Greece; 2009. pp. 1-6. doi: 10.1109/ICDSP.2009.5201143

[33] Sufi F, Khalil I. Faster person identification using compressed ECG in time critical wireless telecardiology applications. Journal of Network and Computer Applications 2011; 34 (1): 282-293. doi: 10.1016/j.jnca.2010.07.004

[34] Tantawi MM, Revett K, Salem AB, Tolba MF. A wavelet feature extraction method for electrocardiogram (ECG)-based biometric recognition. Signal, Image and Video Processing 2015; 9 (6): 1271-1280. doi: 10.1007/s11760-013-0568-5

[35] Tan R, Perkowski M. Toward improving electrocardiogram (ECG) biometric verification using mobile sensors: a two-stage classifier approach. Sensors 2017; 17 (2): 410. doi: 10.3390/s17020410

[36] Coutinho DP, Fred AL, Figueiredo MA. One-lead ECG-based personal identification using Ziv-Merhav cross parsing. In: IEEE 20th International Conference on Pattern Recognition; İstanbul, Turkey; 2010. pp. 3858-3861. doi: 10.1109/ICPR.2010.940

[37] Karimian N, Guo Z, Tehranipoor M, Forte D. Highly reliable key generation from electrocardiogram (ECG). IEEE Transactions on Biomedical Engineering 2016; 64 (6): 1400-1411. doi: 10.1109/TBME.2016.2607020

[38] Odinak I, Lai PH, Kaplan AD, O'Sullivan JA, Sirevaag EJ et al. ECG biometric recognition: a comparative analysis. IEEE Transactions on Information Forensics and Security 2012; 7 (6): 1812-1824. doi: 10.1109/TIFS.2012.2215324

[39] Liu H, Motoda H. Feature Extraction, Construction and Selection: A Data Mining Perspective. Boston, MA, USA: Kluwer Academic, 1998.

[40] Biel L, Pettersson O, Philipson L, Wide P. ECG analysis: a new approach in human identification. IEEE Transactions on Instrumentation and Measurement 2001; 50 (3): 808-812. doi: 10.1109/19.930458

[41] Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Transactions on Biomedical Engineering 1985; 32 (3): 230-236. doi: 10.1109/TBME.1985.325532

[42] Chapra SC, Canale RP. Numerical Methods for Engineers. Boston, MA, USA: McGraw-Hill Higher Education, 2010.

[43] Yu SN, Chou KT. Integration of independent component analysis and neural networks for ECG beat classification. Expert Systems with Applications 2008; 34 (4): 2841-2846. doi: 10.1016/j.eswa.2007.05.006

[44] Jain A, Nandakumar K, Ross A. Score normalization in multimodal biometric systems. Pattern Recognition 2005; 38 (12): 2270-2285. doi: 10.1016/j.patcog.2005.01.012

[45] Zheng G, Wang YR, Qin Q, Li Y, Li ZY. Comparative study of ECG based identification. Applied Mechanics and Materials 2015; 713: 700-703. doi: 10.4028/www.scientific.net/AMM.713-715.700.

[46] Yu SN, Chen YH. Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. Pattern Recognition Letters 2007; 28 (10): 1142-1150. doi: 10.1016/j.patrec.2007.01.017

[47] Safont G, Salazar A, Soriano A, Vergara L. Combination of multiple detectors for EEG based biometric identification/authentication. In: IEEE International Carnahan Conference on Security Technology; Boston, MA, USA; 2012. pp. 230-236. doi: 10.1109/CCST.2012.6393564

[48] Orfanidis SJ. Introduction to Signal Processing. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[49] Abdi H, Edelman B, Valentin D, Dowling WJ. Experimental Design and Analysis for Psychology. Oxford, UK: Oxford University Press, 2009.

[50] Rozza A, Lombardi G, Casiraghi E, Campadelli P. Novel Fisher discriminant classifiers. Pattern Recognition 2012; 45 (10): 3725-3737. doi: 10.1016/j.patcog.2012.03.021.

[51] Luz EJDS, Menotti D, Schwartz WR. Evaluating the use of ECG signal in low frequencies as a biometry. Expert Systems with Applications 2014; 41 (5): 2309-2315. doi: 10.1016/j.eswa.2013.09.028

[52] Irvine JM, Wiederhold BK, Gavshon LW, Israel SA, McGehee SB et al. Heart rate variability: a new biometric for human identification. In: Proceedings of the International Conference on Artificial Intelligence; Las Vegas, NV, USA; 2001. pp. 1106-1111.

[53] Chiu CC, Chuang CM, Hsu CY. A novel personal identity verification approach using a discrete wavelet transform of the ECG signal. In: IEEE International Conference on Multimedia and Ubiquitous Engineering; Busan, South Korea; 2008. pp. 201-206. doi: 10.1109/MUE.2008.67

[54] Fang SC, Chan HL. QRS detection-free electrocardiogram biometrics in the reconstructed phase space. Pattern Recognition Letters 2013; 34 (5): 595-602. doi: 10.1016/j.patrec.2012.11.005

[55] Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. Machine Learning 2002; 46 (1-3): 131-159. doi: 10.1023/A:101245032