



# ECG biometric authentication based on non-fiducial approach using kernel methods



Maryamsadat Hejazi<sup>a,\*</sup>, S.A.R. Al-Haddad<sup>a</sup>, Yashwant Prasad Singh<sup>b</sup>,  
Shaiful Jahari Hashim<sup>a</sup>, Ahmad Fazli Abdul Aziz<sup>c</sup>

<sup>a</sup> Department Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia

<sup>b</sup> Department of Computer Science and Engineering, Manav Rachna College of Engineering, Sector-43, Surajkund Road, Faridabad, India

<sup>c</sup> Faculty of Medicine and Health Sciences, Universiti Putra Malaysia (UPM), 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia

## ARTICLE INFO

### Article history:

Available online 4 March 2016

### Keywords:

ECG biometric recognition  
Non-fiducial feature extraction  
Kernel methods  
Dimensionality reduction  
Gaussian OAA SVM

## ABSTRACT

Identity recognition faces several challenges especially in extracting an individual's unique features from biometric modalities and pattern classifications. Electrocardiogram (ECG) waveforms, for instance, have unique identity properties for human recognition, and their signals are not periodic. At present, in order to generate a significant ECG feature set, non-fiducial methodologies based on an autocorrelation (AC) in conjunction with linear dimension reduction methods are used. This paper proposes a new non-fiducial framework for ECG biometric verification using kernel methods to reduce both high autocorrelation vectors' dimensionality and recognition system after denoising signals of 52 subjects with Discrete Wavelet Transform (DWT). The effects of different dimensionality reduction techniques for use in feature extraction were investigated to evaluate verification performance rates of a multi-class Support Vector Machine (SVM) with the One-Against-All (OAA) approach. The experimental results demonstrated higher test recognition rates of Gaussian OAA SVMs on random unknown ECG data sets with the use of the Kernel Principal Component Analysis (KPCA) as compared to the use of the Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA).

© 2016 Elsevier Inc. All rights reserved.

## 1. Introduction

Biometric recognition technology provides a reliable security system through identity verification and recognition of individuals based on their inherent characteristics in both modes of identification and verification or authentication [1]. The human features can include physiological or behavioural traits. Physiological features include face, iris, and fingerprints whereas keystroke dynamics, gait and voice/speech are behavioural features. However, these modalities are not robust enough against falsification. Some instances of attacks in the biometrics security system involve the use of latex for recreation of a fingerprint, voice imitation and the application of contact lenses copied from original iris features [2].

A new generation of biometric identity recognition modalities has been introduced extensively during the last decades. It includes biosignals which are typically utilised for clinical diagnostic purposes such as Electrocardiogram (ECG) and many others [3,4]. ECG is a non-invasive diagnostic method which is effective, simple

and involves a low-cost procedure to gather information on structural and functional heart muscle activity and other body tissues over time. Each individual has unique ECG characteristic signals such as universality, uniqueness and liveness detection and thus, offer robustness against falsification using fraudulent techniques [5–7]. However, despite these benefits, the technique encounters a number of difficulties due to time-varying nature of the ECG signals, irregular conditions of cardiac disorders, and the long waiting period to collect the ECG data [4,7,8]. Nevertheless, the ECG biometric technology can be widely applied in government and commercial environments, secure public and travel documents, health monitoring systems, forensic systems (including identifying criminals and law enforcement) [9] and distributed systems such as smart cards and many others [2,10,11].

Performance is the most important and significant criteria to evaluate a biometric recognition system [12]. As the ECG is not periodic and highly repetitive signal [2], the ECG biometric systems encounter a number of difficulties in building an accurate identity recognition system. The challenges occur mainly in feature extraction from ECG signals and so in biometric recognitions [4,11,12]. The statistical methods, machine learning and data mining tech-

\* Corresponding author.

E-mail address: Hejazi\_m2002@yahoo.com (M. Hejazi).

niques are suitable for dealing with feature extraction learning and pattern classification (recognition) problems and overcome overfitting and improve generalisation property in numerous real-world and biometric applications [13].

A typical normal heartbeat of ECG wave consists of three main components: P wave, QRS complex and T wave. Fiducial features are amplitudes of P, R and T waves, the temporal distance between wave boundaries (onset and offset of the P, Q, R, S and T waves), the area of the waves, and slope information [11]. The ECG biometric systems based on fiducial detection rely entirely on the localisation of the points in a heartbeat cycle. The fiducial feature extraction method proposed by Biel et al. [3] is among the earliest research in the use of ECG for human identification. The research has showed the impact of a number of features on misclassification and personal identification that used only one lead ECGs. The identification performance of the system was 100% over 20 subjects. Another study based on fiducial features by Israel et al. [14] has proven the effects of an individual's mental stress, and the placement of ECG leads to identification performance. The system achieved 100% subject recognition and 81% heartbeat recognition rate for a population of 29 individuals. Most of the proposed methods in fiducial-based ECG biometric have employed the signals that were acquired using standard leads on the chest and limbs using clinical grade equipment (clinical source data). [15–17] have conducted experiments on standard one-lead ECG signals over a different number of healthy subjects including 25, 76, and 51, respectively. The identification accuracies of these systems were respectively obtained 99, 98, and 99.85%, and equal error rates (EERs) of 1.88% [16] and 0.01% [17] in authentication mode. On the other hand, some studies have proposed fiducial based methodologies which extracted features from characteristic points on ECG recordings acquired at the hands and/or fingers in an off-the-person approach (non-clinical source data) [18–22]. The ECG signals captured from fingers and hands are noisier than the standard ECG signals acquired on the chest using clinical grade equipment, so it leads to make subsequence processing more difficult [18]. However, the fiducial-based algorithms mentioned in the above study and in many other studies have faced the difficulty of missing a number of efficient ECG morphological attributes [4]. In addition, there is no universally acknowledged rule that is able to detect the boundary of fiducial points, so complexity of biometric system is overall increased [7,23].

Many researches based on a non-fiducial approach have been carried out to overcome the fiducial based problems since 2006. The most important advantage of non-fiducial approach over the fiducial ones is in its simplicity of computation as it does not have to determine wave boundaries and fiducial points' detection. Indeed, non-fiducial based techniques extract discriminative information within the ECG waveform without having any particular reference points in heart beat cycles as features. Some methodologies have been presented different types of non-fiducial based features like a sequence of symbols [17], wavelet coefficients [24], and autocorrelation coefficients [23]. In [17], non-fiducial based features were extracted by simple analysis of quantisation to convert a set of ECG waveforms into a set of strings. The identification rate of the system was reported as 99.39% and equal error rate of 0.13% in authentication, tested over 51 healthy subjects. Chan et al. [24] has proposed a human identification system based on collecting ECG signals from fingers across multiple sessions. The Wavelet Distance (WDIST) was used to measure a distance between wavelet coefficients. The system achieved 89% identification rate on a dataset of 50 subjects. Plataniotis et al. [23] have suggested a new and robust approach for feature extraction that relies on a combination of Autocorrelation (AC) and Discrete Cosine Transform (DCT) of the overall morphology of ECG waveform. The DCT has been applied to the autocorrelation (AC) sequence

of windowed ECG data segments for dimension reduction in the non-fiducial approach. The identification performance was 97.8% as tested over standard one-lead ECG signals from 13 subjects. However, the non-fiducial approach based on autocorrelation involves the challenge of high-dimensional data and so analysis or processing is a computationally demanding task, referred to in the literature as the curse of dimensionality [4]. The undesirable template size can have an impact on overall recognition, and it is not able to predict accurately [12]. Thus, it is important to handle data adequately through the construction of a meaningful low-dimensional space form a high-dimensional observation space [25]. Dimensionality reduction can be an effective solution to prevent the curse of dimensionality and enhances the efficiency of learning methods by extracting a number of relevant and significant features often using linear and nonlinear dimensionality reduction techniques [26]. Agraftoti and Hatzinakos [27] dealt with the problem of high-dimensional data in this non-fiducial approach with the use of a combination of AC with Linear Discriminant Analysis (LDA) for human identification system based on fusion of all of the 12 ECG leads' data. The AC/LDA algorithm has been also utilised in studies of [4] and [28] in order to extract discriminative features from one-lead ECG signals collected based on two-session and single-session data (non-clinical source data), respectively. LDA is a linear supervised learning method which has been extensively used in dimension reduction and classification areas for higher inter-class variability [29] in order to control misclassification errors. The linear dimension reduction methods cannot find adequate structures in real-world data that are nonlinear. In recent years, many different techniques have been broadly developed to solve the nonlinear dimensionality reduction problem by using the basic idea of kernel-based learning methods [25,30] such as the Kernel Principal Component Analysis (KPCA) [31]. The main concept of kernel functions is implicit nonlinear mapping of observations in the input space into higher dimensional feature space to be optimally separate input patterns [32]. To the best of our knowledge, there are no reports in the literature about the non-fiducial approach based on an autocorrelation in conjunction with nonlinear dimension reduction methods over non-clinical source data.

The biometric recognition systems in identification and verification modes are classifiers and they are considered respectively as one-to-many and one-to-one problems. The identification systems recognise a subject by comparing an individual's own biometric properties with all enrolled templates in a database to find the highest probability match. The verification systems will then indicate the individual's claimed identity as either accepted or rejected [1]. One verification approach is to compare the probabilities estimate during identification against a threshold to obtain verification decisions. Another approach is to perform a binary classification problem to achieve a probabilistic learning model when the number of individuals in the systems is large [1,33]. The biometric recognition performance is assessed by measuring two misclassification error rates of False Match Rate (FMR) and False Non-match Rate (FNMR) that are respectively termed as False Acceptance Rate (FAR) and False Rejection Rate (FRR) [1]. Training optimal recognition model can be effective on the highest match between the ECG test signals and a collection of training feature vectors. Therefore, the proper selection of robust classification methods can increase generalisation capability and enhance ECG biometric recognition accuracy. General recognition algorithms have been frequently applied in the use of the ECG biometrics including supervised classifiers such as Linear Discriminant Analysis (LDA) [14], K Nearest Neighbours (KNN) [27,6] and also in several research on Support Vector Machines (SVMs) [34,35] and other classification algorithms [11]. SVM is a kernel-based classifier based on a statistical learning theory in machine learning [36] that can attain predictive accuracy with reduced generalisation errors in object detection and classifi-

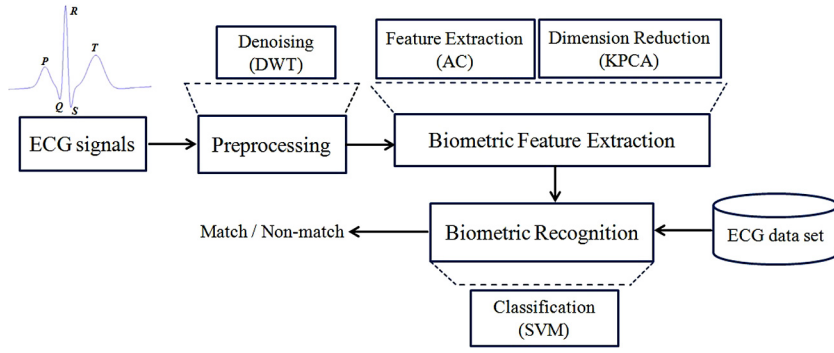


Fig. 1. General scheme of the ECG authentication system proposed.

cation problems [32]. Multi-class SVM classification is a supervised learning problem that has been proposed as an effective extension of binary classification. One-Against-All (OAA) is a multi-class approach which is proposed for biometric verification problem in the current paper by decomposing multi-class SVM into several binary classifiers.

This paper proposes a novel ECG verification framework based on non-fiducial approach using kernel methods in both feature extraction and classification after denoising signals with Discrete Wavelet Transform (DWT). The central consideration of the paper is to evaluate the effect of different linear and nonlinear dimension reduction algorithms in the recognition or authentication performance of a multi-class SVM, and to determine whether the learning classifier is robust in identity verification for individual recognition of unlabelled test data (unknown) by setting training samples.

The paper is organised as follows: Section 2 describes the methodology of the study, relevant technical backgrounds, and an investigation on the systematic analysis of some of the technical backgrounds. Section 3 explains how ECG data sets were collected and how design parameters for multi-class SVMs were chosen. The experimental recognition rates of the proposed feature extraction and recognition system over the ECG data sets are presented in Section 4. Finally, a conclusion and suggestions for future research are discussed in Section 5.

## 2. Methodology

The fundamental concept of underlying biometric technologies involves in different phases of data acquisition, preprocessing, feature extraction and classification. Fig. 1 depicts the general block diagram of the ECG authentication methodology proposed in this research paper. Preprocessing of ECG signals is conducted to remove noise of ECG signals through the filtering and threshold techniques. This is followed by the feature extraction phase to detect non-fiducial features to extract significant feature vectors. In classification phase, the biometric recognition system learns an optimal decision model to best match between training vectors enrolled in ECG dataset and test vectors.

### 2.1. ECG data

Experiments on ECG signals collected were carried out at the BioSec.Lab,<sup>1</sup> at the University of Toronto to evaluate the performance of the proposed methodology. The database contains 68 biometric samples which were captured from 52 healthy volunteers in one session and the remaining ones in two sessions at rest position. ECG data acquisition was done by Vernier EKG-BTA sensor

with a sampling frequency rate of 200 Hz which was attached on the subject's wrists to record one-lead ECG. Every database record contains a 3-minute conventional lead I ECG signal captured from each volunteer [4,10]. In this study, a record from each individual was applied to form the ECG datasets.

### 2.2. Preprocessing

A single normal cycle of the ECG waveform that corresponds to a heartbeat is shown in Fig. 1. Any changes of cardiac function in each heartbeat makes a different deflection on the ECG pulse as series of positive and negative waves consisting of P, QRS complex and T. However, the collected ECG signals are usually corrupted by several sources of noise and therefore, it is not reasonable to perform feature extractions and classifications accurately [2,37]. Many filtering techniques are available for different signal types in order to reduce noise and improve signal-to-noise ratio (SNR) values including digital filters, adaptive denoising methods (involving Least Mean Squares and Recursive Least Squares) and time frequency domain representation (such as wavelet transform thresholding techniques). ECG is a non-stationary signal, so time-frequency representation methods are typically suited for its signal processing than the above mentioned adaptive denoising methods. Similarly, wavelet transform is a non-stationary signal processing technique that can be applied to sub-band coding data compression, feature extraction and denoising a signal without appreciable degradation [18,37–40]. Indeed, the wavelet transform can be thought of as an extension of Short Time Fourier Transform (STFT) that uses variable window sizes, depending on the frequency range of analysis. The wavelet transform can be classified into continuous and discrete. The Continuous Wavelet Transform (CWT) describes how a signal function  $f(t)$  is decomposed over entire time by fundamental mother wavelet  $\psi$  into a set of scale and position variables as shown in Eq. (1):

$$\text{CWT}(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$$

$$a \in \mathbb{R}^+ - \{0\}, b \in \mathbb{R} \quad (1)$$

where 'a' is a positive real number named as dilation or contraction parameter, and 'b' is a real number known as window translation or position parameter. The CWT method transforms a continual signal into highly redundant wavelet coefficients. Discrete Wavelet Transform (DWT) starts with a discrete set of data and considers a dyadic set of scales (power of two). The DWT is used to eliminate the difficulty of redundancy in CWT. DWT gives both time and frequency information through decomposition of a signal into sets of approximation (low frequency part of initial signal) and detail (high frequency part of initial signal) based on powers of two in different levels as shown in Eq. (2):

<sup>1</sup> <http://www.comm.utoronto.ca/~biometrics/>.

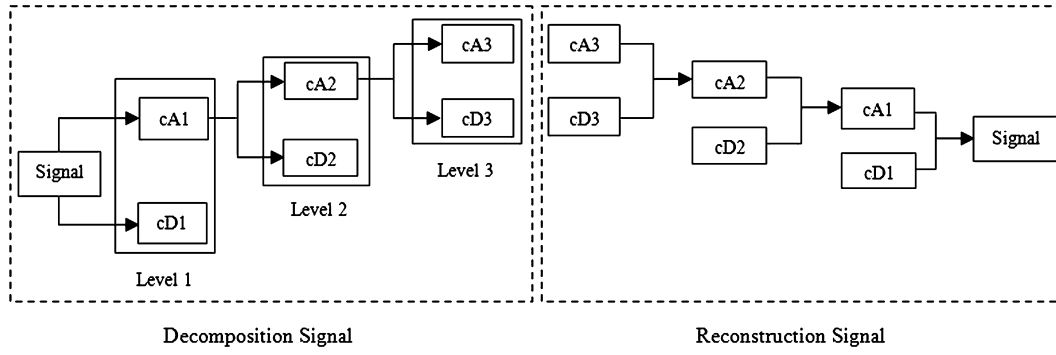


Fig. 2. A scheme of discrete wavelet decomposition and reconstruction.

$$\text{DWT}(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$$

$$a = 2^j, \quad b = k2^j, \quad k, j \in \mathbb{Z} \quad (2)$$

where  $\mathbb{Z}$  is a set of integers. Wavelet based denoising procedure consists of three steps: 1) choosing a mother wavelet and a decomposition level to perform the signal denoising through wavelet deconstruction, 2) performing appropriate soft or hard thresholding methods in wavelet domain, and 3) reconstructing the signal with the modified DWT coefficients via inverse wavelet transform. Fig. 2 shows a general scheme of the discrete wavelet decomposition and reconstruction process at three levels. The proper wavelet basis function leads to the maximisation of cross-correlation between coefficient values in the wavelet domain and the original signal, and it can be more effective in thresholding of wavelet coefficients for denoising of original signal [39]. The appropriate mother wavelet selection and the best decomposition level can be evaluated statistically based on measuring criteria including the higher Cross-Correlation (CC) coefficients, lower Root Mean Square Error (RMSE), higher Retained Energy (REnergy), highest Signal-to-Noise Ratio (SNR) and lowest Reconstruction Error (RError) between the original ECG signals and the wavelet coefficients at a particular level. The criteria are described below in Eqs. (3)–(5):

$$\text{CC}_{s_0 s_r}(\tau) = \sum_{n=0}^{N-|\tau|-1} s_0(n) s_r(n + \tau) \quad (3)$$

where  $\tau \ll N$  is time lag

$$\text{RMSE} = \sqrt{\frac{\sum_{n=0}^N (s_0(n) - s_r(n))^2}{N}} \quad (4)$$

$$\text{SNR} = \log\left(\frac{\sum_{n=0}^N s_r^2(n)}{\sum_{n=0}^N (s_0(n) - s_r(n))^2}\right) \quad (5)$$

where  $s_0$  is the original signal sample,  $s_r$  is the reconstructed signal at each level, and  $N$  is the length of signal. Moreover, the shape of the scaling function of wavelet bases can also have an impact on the selection of mother wavelet. Daubechies, Symlet and Coiflet are orthogonal and compact support mother wavelets for denoising of ECG signals that have been used previously [37,39]. Therefore, this paper extends these mother wavelet functions.

Table 1 depicts the average denoising influence of different discrete wavelet transforms in the two optimal levels of decomposition among all tested levels on the 52 ECG signals. It can be seen from Table 1 that the performance criteria for wavelet denoising functions is very close in both decomposition levels. However, the performance of SNR and RError properties in db7 and coif3 at level 4 is better than the performance of other wavelet methods. Hence, these mother wavelets at level 4 can have an impressive performance for denoising the ECG signals.

Table 1

ECG denoising performance of wavelet transforms on different decomposition levels.

Level	MW*	CC	RMSE	REnergy	SNR	RError
4	db4	0.9920	0.1296	97.23	13.03	2.20
	<b>db7</b>	<b>0.9920</b>	<b>0.1294</b>	<b>97.24</b>	<b>13.04</b>	<b>2.19</b>
	sym8	0.9921	0.1292	97.25	12.97	2.27
	sym9	0.9921	0.1292	97.25	12.97	2.28
	<b>coif3</b>	<b>0.9920</b>	<b>0.1294</b>	<b>97.24</b>	<b>13.04</b>	<b>2.19</b>
	coif5	0.9921	0.1292	97.25	12.98	2.27
5	db4	0.9901	0.1446	96.86	12.72	2.57
	db7	0.9901	0.1448	96.87	12.86	2.40
	sym8	0.9901	0.1449	96.87	12.86	2.40
	sym9	0.9901	0.1449	96.87	12.78	2.49
	coif3	0.9901	0.1448	96.87	12.86	2.39
	coif5	0.9901	0.1450	96.88	12.88	2.37

\* Mother wavelet.

Wavelet denoising entails soft or hard thresholding methods in which significant wavelet coefficients are selected by applying a specific threshold value ( $\lambda$ ). Hard thresholding sets coefficients below the defined  $\lambda$  to zero and helps to remove noisy data without any change in the main characteristics of the original signals. Soft thresholding, on the contrary, sets coefficients below the value  $\lambda$  to zero and also shrinks other coefficients according to  $\lambda$  [41]. Therefore, the selection of the threshold values is able to overcome the challenge of having too much noise reduction and the change in signal shape as these are important criteria of a reliable signal threshold for denoising. Threshold selection methods estimate a different value of  $\lambda$  with or without rescaling, so that the techniques are able to have a direct influence over the quality of output denoised signal. Minimax rule utilises precomputed thresholds to minimise a constant term in the upper bound for the minimax risk of estimating a function by using a threshold estimator. The first technique uses Sqrtwolog rule that has been proposed as an alternative to minimax threshold to calculate the threshold values  $\lambda$  with the use of universal threshold method  $\lambda_n^u = (2 \log n)^{1/2}$  where  $n$  is the length of signal [42]. The second is Rigrsure rule which is an adaptive thresholding method that uses Stein Unbiased Risk Estimator (SURE) principle to minimise error. The threshold value  $\lambda$  is estimated to  $\lambda = \sigma \sqrt{w_b}$  where  $w_b$  is bth coefficient at minimum risk from vector  $w$  and  $\sigma$  is the standard deviation of noisy signal. The third is Heursure rule which is a combination of the Sqrtwolog and Rigrsure threshold techniques. If the signal-to-ratio is very small, the Sqrtwolog threshold method gives better estimation than SURE technique. A small threshold yields noisy results, but large thresholds can remove significant segments of signals and miss the significant details of the signals. Penalisation rules are a type of Birgé–Massart strategy that uses level dependent thresholds in wavelet coefficients selection rule  $n_j = \frac{m}{(j+2-i)^\alpha}$  where  $j$  is decomposition level,  $m$  is the length of approximation coefficients over 2 and  $\alpha > 1$  [43].



**Table 2**  
Selection of proper wavelet function for the ECGs' denoising.

Wavelet	Threshold rules	Soft/hard method	SNR	RMSE	PRD	PERFL2	PERF
db7	Rigrsure	Soft	5.14	0.07	0.75	99.49	63.54
	Heursure		5.09	0.07	0.78	99.49	66.56
	Sqtwolog		4.59	0.10	0.99	99.48	94.18
	Minimaxi		4.75	0.09	0.92	99.48	89.27
	Penalhi <sup>*</sup>		4.78	0.07	0.90	99.48	63.61
	Penalme <sup>**</sup>		4.84	0.07	0.88	99.48	60.57
	Rigrsure	Hard	5.26	0.06	0.71	99.49	63.54
	Heursure		5.19	0.06	0.74	99.49	66.56
	Sqtwolog		4.83	0.08	0.88	99.49	94.18
	Minimaxi		5.01	0.07	0.81	99.49	89.27
	Penalhi <sup>*</sup>		4.96	0.07	0.83	99.49	63.61
	Penalme <sup>**</sup>		5.01	0.06	0.81	99.49	60.57
Coif3	Rigrsure	Soft	5.21	0.06	0.73	99.32	57.54
	Heursure		5.17	0.06	0.75	99.31	58.95
	Sqtwolog		4.76	0.09	0.91	99.31	91.65
	Minimaxi		4.91	0.08	0.85	99.31	87.50
	Penalhi <sup>*</sup>		4.84	0.07	0.88	99.31	66.58
	Penalme <sup>**</sup>		4.89	0.07	0.86	99.31	63.62
	<b>Rigrsure</b>	<b>Hard</b>	<b>5.30</b>	<b>0.06</b>	<b>0.70</b>	<b>99.32</b>	<b>57.54</b>
	Heursure		5.24	0.06	0.72	99.32	58.95
	Sqtwolog		5.01	0.07	0.81	99.32	91.65
	Minimaxi		5.13	0.07	0.76	99.32	87.50
	Penalhi <sup>*</sup>		5.02	0.06	0.80	99.32	66.58
	Penalme <sup>**</sup>		5.06	0.06	0.79	99.32	63.62

<sup>\*</sup>  $\alpha = 3$ .

<sup>\*\*</sup>  $\alpha = 1.5$ .

Many different analysis criteria are applied in order to measure similarities and difference between the thresholding wavelet coefficients obtained at each decomposition level and the original ECG signals. The selection of proper soft or hard threshold methods is based on the criteria including the highest Signal-to-Noise Ratio (SNR), lowest Root Mean Square Error (RMSE), lowest percentage root mean square difference or PRD, highest L2 norm recovery value (PERFL2) and lowest percentage of zero coefficients used during reconstruction (PERF) that are defined with:

$$PRD = \sqrt{\frac{\sum_{n=0}^N (s(n)_{\text{original signal}} - s(n)_{\text{reconstructed signal}})^2}{\sum_{n=0}^N (s(n)_{\text{original signal}})^2}} \quad (6)$$

$$PERF = \frac{\text{Number of zero coefficients at current level}}{\text{Number of coefficients}} \quad (7)$$

$$PERFL2 = \frac{(\text{vector} - \text{norm}(\text{coeffs of the current decomposition}, 2))^2}{(\text{vector} - \text{norm}(\text{original signal}, 2))^2} \quad (8)$$

Table 2 demonstrates the average denoising performance of the properties over the 52 ECG signals on six wavelet thresholding rules with hard and soft thresholding methods based on db7 and coif3 wavelet functions at the optimal decomposition level 4. Table 2 indicates that the Rigrsure rule for both soft and hard methods of coif3 wavelet gives the best results based on the performance of SNR and PRD. Based on the PERF value, Rigrsure of coif3 wavelet has performed well over other thresholding rules in db7 and coif3 wavelets. The major observation from the above results is that coif3 wavelet with Rigrsure rule of hard thresholding gives the best performance for the denoising of the ECG signals. Fig. 3 shows the comparison of a segment of original normal ECG signal (with cyan colour) and the signal denoised (with black colour) using coif3 wavelet at level 4 decomposition.

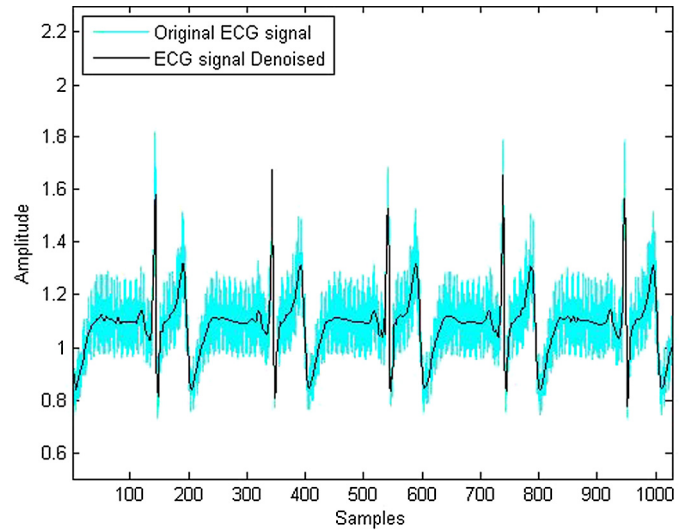


Fig. 3. ECG signal denoising with coif3 at level 4.

### 2.3. ECG biometric feature extraction

This paper focuses on ECG biometric feature extraction encompassing a combination of the non-fiducial based approach and different types of linear and nonlinear dimensional reduction techniques. Each approach is briefly described below.

#### 2.3.1. Non-fiducial feature extraction

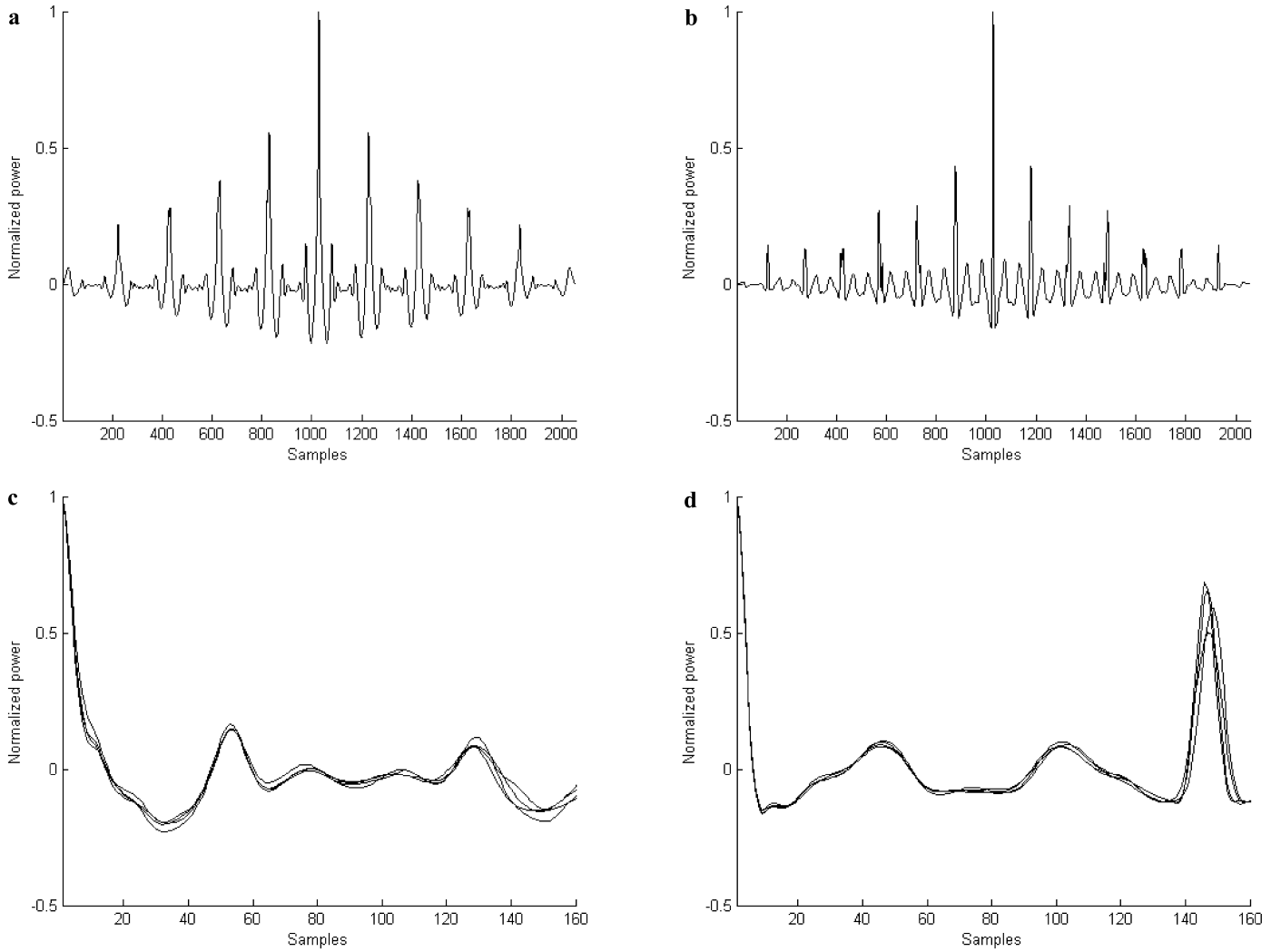
The ECG is not a stationary signal and it consists of repetitive waveform patterns. The motivation behind non-fiducial approach is the use of normalised autocorrelation (AC) method on non-overlapping windows of the denoised individual ECG signal for detecting nonrandom and discriminative pattern without the use of fiducial point's boundaries [23]. Indeed, autocorrelation gives an automatic shift invariant feature set that represents repetitive characteristics over multiple heartbeat cycles [5]. Thus, the autocorrelation of ECG windows can be an effective method in identity verification process [6]. The autocorrelation coefficients  $\hat{R}_{xx}(m)$  are computed by Eq. (9):

$$\hat{R}_{xx}(m) = \frac{\sum_{i=0}^{N-|m|-1} x(i)x(i+m)}{\hat{R}_{xx}(0)} \quad (9)$$

where  $\hat{R}_{xx}(m)$  is the correlation of an ECG window  $x(i)$  and  $x(i+m)$  that is shifted by a time lag  $m = 0, 1, \dots, M-1$ ;  $M \ll N$ , and  $N$  is the length of number of data points inside an ECG window  $x(i)$ . Large variations in amplitude are among the pulses of the same subject, so normalisation is essentially required. Normalisation and control of bias factor are carried out by dividing  $\hat{R}_{xx}$  with the maximum value  $\hat{R}_{xx}(0)$ . Experiments conducted in this study for feature extraction based on the autocorrelation method show that the ECG signals have been windowed using a non-overlapping rectangular window of length 1028 coefficients which is equal to 5135 msec duration. The normalised AC is computed for each window by Eq. (9). A segment beginning at zero lag of the autocorrelation sequence is only preserved in order to do more analysis. The length of this segment is about equal to the duration of a QRS pulse with less variability in time [6]. Fig. 4 shows normalised autocorrelation sequence of an ECG window for two subjects, and also zooms in to the number of AC coefficients from different windows of both subjects.

#### 2.3.2. Dimension reduction

The extracted autocorrelation coefficients of each ECG window can be directly utilised for classification, but the dimension-



**Fig. 4.** The normalised autocorrelation sequence of subjects A (a) and B (b). Zoom in to 160 AC coefficients of both subjects A (c) and B (d) from different ECG windows.

ality of an AC vector is significantly high making it unsuitable for cost efficient systems and controlling false rejection rate [6]. Thus, dimension reduction is required to solve this problem as it converts a high-dimensional input space into a meaningful low-dimensional output space using the broadly linear and nonlinear techniques. The Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are popular linear statistical learning techniques which have been widely used in computer vision, biometrics and other different areas [25]. The computation and memory requirement of PCA are smaller than LDA, so it is proper for a real time system [6]. However, the linear methods may fail to find essential data structure of nonlinear patterns in input space. Kernel-based techniques have been developed to handle nonlinear dimension reduction problem by mapping implicitly input observations into high dimensional feature space. Kernel-based methods are typically used in many real-world applications including Kernel Principal Component Analysis (KPCA) and Kernel Linear Discriminant Analysis (KLDA). The paper aims to evaluate the effects of the dimensionality reduction methods (LDA, PCA and KPCA) on ECG verification performance to determine whether the nonlinear method is more suitable for dimensionality reduction of the autocorrelation coefficients of ECG windows compared to the linear methods. Mathematical concepts for every one of the stated methods are given briefly in the following sections.

**2.3.2.1. Linear discriminant analysis (LDA)** LDA is a statistical supervised learning method for classification and dimensionality reduction. A given data set consists of a number of  $n$  observation  $X = \{(x_i, y_i)\}_{i=1}^n \in \mathbb{R}^{n \times d}$ , where  $y_i \in \{1, 2, \dots, k\}$  denotes the corresponding class label of  $k$  classes, and  $x_i \in \mathbb{R}^d$  indicates the  $d$ -dimensional input data. Let the data set is defined as  $X = \{X_i\}_{i=1}^k$  and  $X_i \in \mathbb{R}^{n_i \times d}$  is the data matrix of the  $i$ th class consisting of the number of samples  $n_i$  and so  $\sum_{i=1}^k n_i = n$  (total samples of all classes). LDA computes a linear transformation  $\mathbf{w} \in \mathbb{R}^{d \times l}$  that transforms  $x_i \in \mathbb{R}^d$  in the  $d$ -dimensional space to a vector  $x_i^l = \mathbf{w}^T x_i \in \mathbb{R}^l$  in the  $l$ -dimensional space ( $l < d$ ) [44]. An optimal linear transformation is computed by simultaneously finding minimum intra-class distance and maximum inter-class distance so that it attains maximum class discrimination. The set of  $l$  feature basis vectors  $\{\mathbf{w}_m\}_{m=1}^l$  can be approximated by maximising the following Fisher's ratio function  $J(\mathbf{w})$  as shown in Eq. (10):

$$J(\mathbf{w}) = \arg \max_{\mathbf{w}} \frac{|\mathbf{w}^T S_B \mathbf{w}|}{|\mathbf{w}^T S_W \mathbf{w}|} \quad (10)$$

The vector  $\mathbf{w}$  that maximises  $J(\mathbf{w})$  satisfies the following condition and is known as a generalised eigenvalue problem in Eq. (11) [6,29,44,45]:

$$(S_W)^{-1} S_B \mathbf{w}_i = \lambda_i \mathbf{w}_i \quad (11)$$

where  $\mathbf{w} = [\mathbf{w}_1, \dots, \mathbf{w}_l]$  and  $S_B$  and  $S_W$  are between-class and within-class variance matrices which are defined by Eq. (12) and Eq. (13) [44]:

$$S_B = \frac{1}{n} \sum_{j=1}^k (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^T \quad (12)$$

$$S_W = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{\mathbf{x}}_j)(x_{ij} - \bar{\mathbf{x}}_j)^T \quad (13)$$

where  $x_{ij}$  is the  $i$ th sample of  $j$ th class label, and  $\bar{\mathbf{x}}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$  is the mean of set of samples of each class and  $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n x_i$  is the global mean. However, classical LDA needs the so-called total scatter matrices (as:  $S_T = S_B + S_W$ ) to be nonsingular if dimensionality of data is larger than the sample size which the case for many high-dimensional and low samples size data [44]. Maximisation of Fisher's ratio requires large separations between samples of different classes, and small variance between samples of each class. The optimal solution to the optimisation problem in Eq. (10) can be achieved by solving the generalised eigenvalue problem defined in Eq. (11). LDA finds the vector  $\mathbf{w}$  as the  $l$  most significant eigenvectors of  $(S_W)^{-1}S_B$  corresponding to the first  $l$  largest eigenvalues  $\lambda \geq 0$  [6]. Indeed, LDA reduces the dimensionality of feature space for a  $k$  class problem with the eigenvectors corresponding to the  $k-1$  largest eigenvalues [44]. The LDA linear projection for a data vector  $\mathbf{x} \in \mathbb{R}^d$  can be derived generally using Eq. (14) before the classification phase, where  $y \in \mathbb{R}^l$  such that  $1 \leq l \leq k-1$ :

$$y = \mathbf{w}^T \mathbf{x} \quad (14)$$

**2.3.2.2. Principal component analysis (PCA)** PCA is a powerful unsupervised learning method to extract significant features based on principal components from a high-dimensional data set by solving an eigenvalue problem. PCA finds directions that have the least reconstruction error by describing as much variance of the data matrix as possible with a set of orthogonal directions [29,46]. Let average  $\bar{\mathbf{x}}$  is subtracted from the unlabelled data set  $X = \{x_i\}_{i=1}^n \in \mathbb{R}^d$ , which gives the mean centred data (zero-mean data). These mean-centred data are used to compute a covariance matrix. PCA diagonalises the covariance matrix  $S_{cov}$  of the multivariate data set  $X$  as in Eq. (15):

$$S_{cov} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{\mathbf{x}})(x_i - \bar{\mathbf{x}})^T \quad (15)$$

where  $n$  is the number of vectors in the data set. The covariance between two vectors reveals the extent in which the patterns are correlated. A transformation of the samples can be computed by using eigenvalue analysis of the covariance matrix  $S_{cov}$  as in Eq. (16):

$$\lambda_i \mathbf{w}_i = S_{cov} \mathbf{w}_i \quad (16)$$

where eigenvalue  $\lambda \geq 0$  and  $\mathbf{w} \in \mathbb{R}^d \setminus \{0\}$  is the nonzero vector  $\mathbf{w} = [\mathbf{w}_1, \dots, \mathbf{w}_d]$  corresponding to eigenvalue  $\lambda$  [46].  $k \leq d$  eigenvectors corresponding to the highest eigenvalues are actually needed to produce a complete basis for the linear transformation. Performance of classifier varies as the number of principal components (features) changes. The dimension of data vector  $\mathbf{x}$  is generally reduced using PCA with the following Eq. (17), where  $y$  is  $k$  dimension, and  $\mathbf{w} = [\mathbf{w}_1, \dots, \mathbf{w}_k]$ :

$$y = \mathbf{w}^T (\mathbf{x} - \bar{\mathbf{x}}) \quad (17)$$

**2.3.2.3. Kernel principal component analysis (KPCA)** Kernel PCA is known as nonlinear technique since PCA is done on nonlinearly mapped data set of input space. Nonlinear techniques refer to the methodology used to attain a minimal generalisation error for data that are nonlinear in input space. KPCA overcomes the limitation of linear methods by nonlinearly mapping input space to a high dimension feature space. Indeed, KPCA is a nonlinear extension of PCA which computes principal components in higher dimension feature space. Kernel-based learning methods use an implicit mapping of the input data into a high dimensional feature space defined by a kernel function, i.e., a function returning the inner product  $(\Phi(x), \Phi(y))$  between the images of two data points  $x, y$  in the feature space. The kernel based methods compute inner products in the feature space directly from the inputs without explicitly transforming  $x$  and  $y$  into the feature space using function  $\Phi$ . Kernel function is known as a kernel trick with the representation form  $K(x_i, x_j) = (\Phi(x_i), \Phi(x_j))$  which allows computing the dot product into Reproducing Kernel Hilbert Space (RKHS) without doing computation of the map  $\Phi$  [32,45,46].

The common kernel functions are given below:

- Polynomial:  $K(x_i, x_j) = ((x_i \cdot x_j) + r)^d$  where  $d \in \mathbb{N}$ ,  $r \in \mathbb{R}$
- Gaussian RBF:  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$  where  $\gamma \in \mathbb{R}$

Obtaining KPCA components of the unlabelled data set  $X = \{x_i\}_{i=1}^n \in \mathbb{R}^d$  only requires solving the same linear eigenvalue problem as described before for linear PCA, but dot product kernel matrix  $n \times n$  (Gram Matrix) instead of covariance matrix of dimension  $d \times d$  problem in PCA. A kernel function  $\Phi$  is used to map centred input vector  $X_i$  from input space,  $\mathbb{R}^d$ , to high dimension feature space  $(F)$ ,  $\mathbb{R}^F$ , as  $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^F$ ,  $F \gg d$ , and  $X_i \rightarrow \Phi(X_i)$ . The covariance matrix is calculated in feature space with dimension  $F \times F$ , which is very high, so the eigenvalues and corresponding eigenvectors are not calculated from covariance matrix. Therefore, the diagonalising Gram  $K = \{K_{i,j}(x_i, x_j)\}$  needs to be centred in feature space and  $F$  must be substituted for nonlinear PCA to capture the nonlinear principal component [45].

Fig. 5 indicates the effect of PCA and RBF KPCA with  $\gamma = 1$  into inter-subject variability for the two persons' normalised AC coefficients. We use a criterion by taking the  $m$  optimal number of components in PCA and KPCA that satisfies approximately  $\sum_{i=1}^m \lambda_i / \sum_{i=1}^d \lambda_i \geq 99$  and 80% respectively, where  $\lambda_i$  is the eigenvalue and  $d$  is the dimensionality of feature space in each method.

## 2.4. ECG biometric recognition

Biometric verification problem is applied to decide whether to accept or reject a user's claimed identity according to the different biometric sample vectors enrolled. Indeed, verification is a binary classification problem due to enrolment of samples of an individual for verification as one class and all other enrolment samples of the remaining persons as other classes [47]. Therefore, different subjects in the ECG data set need different binary classifiers to be recognised by the system. Support Vector Machine (SVM) is a very robust nonparametric classification technique that is based on statistical learning theory in machine learning to deal with problems of classification and regression [36]. The basic theories of SVM have been initially proposed for binary classification problems to increase generalisation ability in order to improve predictive accuracy. The binary SVM algorithms are based on learning a decision boundary with a maximum margin between specified classes either in data space or in feature space for linear and nonlinear SVMs. Nonlinear classification arises whenever the training data points  $(x_i, y_i)$ ,  $x_i \in \mathbb{R}^d$  and class label  $y_i \in \{+1, -1\}$  cannot be separated with a linear classifier in the  $d$ -dimensional input space with lower susceptibility to overfitting.  $K(x_i, x_j)$  defined by

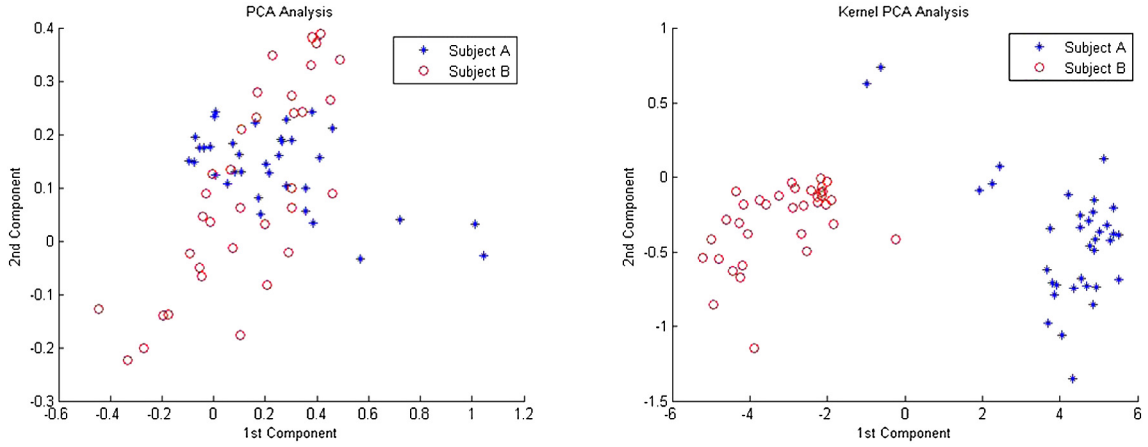


Fig. 5. Effect of PCA (left side) and KPCA (right side) analysis on inter-subject variability of the autocorrelated ECG coefficients of two subjects.

$K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j))$  and is known as the kernel function, which captures the similarity between points  $x_i$  and  $x_j$ . Kernel method is employed in SVM design to construct an optimal linear classifier in feature space which results into nonlinear decision boundary in input space. This margin maximisation is closely associated to the generalisation property of SVMs. SVM is a supervised classifier that has generalisation ability in the sense that it can classify an unseen pattern correctly. There is a trade-off between the size of margin and the number of training errors in SVM methods for the linear decision boundary [48]. Therefore, the learning of optimal decision boundary has an influence on minimising the generalisation errors and performs well on the test data [32].

Many real-world problems such as biometric recognition [47] encounter input data which consists of more than two classes and represented as a set of  $k$  classes  $Y = \{y_1, y_2, \dots, y_k\}$  [32]. Multi-class classification is known as supervised learning problem that extends the binary classification methods. The computational effort and complexity of multi-class classification are more than binary classifiers because of the high connection among a number of different classes. Several approaches have been developed to handle multi-class problems by SVM through their decomposition into  $k$  binary problems. One-Against-All (OAA) approach [48] is the most common multi-class technique which is able to decompose multi-class problem into  $k$  binary classification problems. With regard to this approach, a training vector would be classified under a certain class if the class of SVM accepts data; otherwise, it would be rejected [32,49]. Test data is classified by combining the predictions learned with the binary methods according to OAA approach. A voting strategy is generally utilised to combine the predictions, where the class that gets the highest number of votes is assigned to the test sample. Another possible scheme is to map the binary classifiers output into probability approximates before assigning the test instances to the class with the highest one [32,50]. In this paper, we propose an ECG-based recognition model relying on multi-class SVM classification through OAA approach by using an appropriate kernel function. C-Support Vector Classification (C-SVC) as binary SVM classifier is proposed to build  $k$  binary models for multi-class recognition problems. The mathematical concept of C-SVC is briefly described below:

C-Support vector classification (C-SVC) is a type of binary classification method that has been formulated as the soft-margin SVM primal constrained optimisation problem in the following: the primal constrained optimisation problem for two classes with  $\{x_i\}_{i=1}^n \in \mathbb{R}^d$  and  $y_i \in \{+1, -1\}$  is defined in Eq. (18):

$$\min_{w, b, \xi} \left[ \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^n \xi_i \right] \quad \text{subject to} \quad y_i(w \cdot \Phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, \quad \forall i \in n \quad (18)$$

here,  $w$  is the weight vector,  $b$  is bias term or offset of the hyperplane from the origin,  $\xi_i$  is the margin violation of the  $i$ th training data and  $\Phi(x) : \mathbb{R}^n \rightarrow \mathbb{R}^N$  is a (possibly non-linear) feature mapping that maps a point in  $\mathbb{R}^n$  to a higher dimensional space  $\mathbb{R}^N$ . The first term in the above problem seeks to maximise the margin, while the second term corresponds to empirical risk minimisation.  $C$  is a cost parameter that controls trade-off between maximising the margin size and the minimising the training error. Slack variables ( $\xi_i$ ) demonstrate the estimation of the number of training errors. If  $\xi_i = 0$ , then there is no margin error; otherwise, all slack variables show margin errors. The above quadratic programming problem is computed by using the Lagrange multipliers method by redefining the objective function by attaching constraints  $f(x) = w \cdot \Phi(x) + b$  using Lagrange multipliers into the objective function (known as the primal problem) as an unconstrained optimisation problem. The primal problem can be simplified by reformulating the above primal problem as a dual problem through the Lagrange multipliers (e.g.,  $\lambda_i \geq 0$ ) [32]. The Lagrange multiplier  $\lambda_i$  must be zero unless the training data  $x_i$  satisfies in the constraint  $y_i(w \cdot x_i + b) = 1$ . Such training sample, with  $\lambda_i \geq 0$ , locates on the hyperplanes  $w \cdot x_i + b = \pm 1$  and is known as support vector. The corresponding dual problem for the above constrained optimisation problem of Eq. (18) is the following:

$$\min_{\lambda} \left[ \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i,j=1}^n \lambda_i \lambda_j y_i y_j \Phi(x_i) \cdot \Phi(x_j) \right] \quad \text{subject to} \quad 0 \leq \lambda_i \leq \frac{C}{n}, \sum_{i=1}^n \lambda_i y_i = 0, \quad C > 0, \quad \forall i \in n \quad (19)$$

here,  $\lambda_i$  is the dual variable associated with the  $i$ th primal constraint, and  $\Phi(x_i) \cdot \Phi(x_j)$  is elements of the kernel function, which captures the similarity between points  $x_i$  and  $x_j$ . The primal variable  $w$  is related to the dual variable  $\lambda$  as Eq. (20):

$$w = \sum_{i=1}^n \lambda_i y_i \Phi(x_i) \quad (20)$$

The decision function can be derived from the dual optimisation problem can be represented by Eq. (21):

$$f(x) = \text{sign} \left( \sum_{i,j=1}^n y_i \lambda_i k(x_i, x) + b \right) \quad (21)$$



The bias term  $b$  can be calculated for data points by averaging as shown in Eq. (22):

$$b = f(x) - \sum_{i=1}^n y_i \lambda_i k(x_i, x) \quad (22)$$

The goal of the classification problem is learning an optimal separable hyperplane known as a decision function as defined in Eqs. (21) and (22).

The interpretation of Eq. (21) is that each training point  $x_i$  in the training set is assigned an importance weight of  $\lambda_i$ . The label of a test point  $x$  is decided by a linear combination of a labels of important points similar to  $x$ . It may be noted that this interpretation is consistent with the fact non-support vectors are assigned  $\lambda_i = 0$ , and they are indeed unimportant for decision, since their removal does not affect the decision boundary.

### 3. Experiments

This section evaluates recognition performance comparison of multi-class SVM models based on kernel methods with OAA approach over optimal feature sets which are extracted by using different dimension reduction algorithms described in Section 2.3.2. The objectives of the experiments are to find which linear and nonlinear feature extractors that can be more efficient for ECG biometric verification, and also determine the effect of kernel methods in recognition performance.

With regard to Section 2.3.1, an optimal number of ECG windows with an identical length examined for each subject's ECG recording is 35, so the total windows of the 52 ECG signals (single-session) are 1820. The dimension reduction experiments are carried out on the generic dataset including the 1820 AC windows with the vector of dimension 160 which is the number of normalised AC coefficients for every window. To improve performance of SVM classification algorithms, it is desirable to have data set with proper feature space dimension. For this purpose, in Section 2.3.2, LDA, PCA and KPCA have mathematically been discussed to reduce dimensionality of AC segments and to find significant features in order to enhance recognition performance. The kernel selection has a crucial effect on the SVM classification performance [45]. There are no definite rules for a specific kernel choice except satisfactory performance by simulation study. RBF kernel function with different parameter settings has been experimented in this study for both kernel-based learning methods of nonlinear feature extractor and classification (recognition).

When the supervised LDA method is used to find the significant features by solving the eigenvalue problem, the generic data set must be labelled by taking a class label for a set of AC windows of a subject. In this experiment, the generic data set is labelled by a set of 52 class labels which corresponds to the number of subjects considered in this study. As indicated earlier in Section 2.3.2, LDA can reduce the dimensionality feature space of a  $k$  class problem to  $k - 1$  because the rank of inter-class matrix cannot go beyond than  $k - 1$ . Therefore, the ECG data set is created by the 51 largest eigenvalues that can be a significant for LDA domain to reduce dimensionality of the generic data set. The dimensionality of the generic data set is reduced from 160 to 31 using PCA method which retains 99% of the data's variance. Using KPCA method with RBF kernel in different  $\gamma$  values of 0.01, 0.1, and 1 helps to reduce dimensionality of the generic data set from 160 coefficients to data sets including 54, 107 and 111 features respectively which preserve approximately 99, 98 and 80% of the data's variance. To build OAA SVM multi-class model with OAA approach, the same above set of 52 class labels have to be assigned to vectors of all the ECG data sets which are created by using PCA and KPCA feature extractors.

For each of the ECG data sets predefined, the generalised accuracies of multi-class SVM models are estimated using different values of the RBF kernel parameter  $\gamma$  in  $\{1, 5, 10, 15\}$  and the regularisation constant  $C$  of C-SVC binary classifier in  $\{10, 60, 100\}$ . Then, 3 random data sets are generated from every ECG data sets predefined for each pair of  $(C, \gamma)$ . Therefore, each of the ECG data sets is tested on  $4 \times 3 \times 3 = 36$  combinations. Both training and testing sets are available for the ECG data sets, whereby each is randomly split into the training and test sets (mostly  $\approx 60\%$ :  $40\%$ )-called Percentage Split or Holdout method [32,45]. The number of training vectors is 1092 samples (AC windows) while the other 728 form the unlabelled (unknown) test data sets for classification. The result of the Gaussian OAA SVMs on the training and testing ECG data sets including the means and standard deviations are summarised.

Nevertheless, the proposed framework must be validated by performing it on a set of ECG signals collected from 16 subjects in two recording sessions with regarding Section 2.1, as well as running 10-fold cross-validation technique on each of the ECG data sets predefined from one and two recording sessions.

In this ECG based biometric system, the recognition performance of OAA multi-class SVM models is measured according to True Match Rate (TMR), and the two types of errors of False Non-match Rate (FNMR) and False Match Rate (FMR), also known as classification error rates. TMR measures how well the system is able to correctly match mate samples (classified), and it is known as Window Recognition Rate (WRR) in this verification system. The Subject Recognition Rate (SRR) can also be computed by measuring TMR. The subject recognition rate is a measurement for a number of subjects whose all test samples can be correctly classified into their exact class based on their learning patterns. FNMR refers to the number of mate samples falsely declared as nonmatch (unclassified) whereas FMR refers to the number of nonmate samples recognised as a correct match (misclassified). When the system becomes more secure, FMR should be less than FNMR.

### 4. Results and discussions

This section provides experimental results that consider the verification performance comparison of Gaussian (RBF) OAA SVM classifiers on the ECG data sets predefined by using several different extraction algorithms (combination of autocorrelation and dimension reduction methods), and also investigates the effect of different parameters settings  $(C, \gamma)$  in verification recognition system. The different ECG data sets are displayed in the following tables of results by Dimension Reduction (DR) method and the number of optimal features used (#F) for each of them.

The experimental results given in Tables 3, 4, 5 and 6 indicate that the Gaussian OAA SVM models are very sensitive to the RBF kernel parameter  $\gamma$  than the cost parameter  $C$  for unlabelled test data sets. The parameter  $C$  controls the trade-off between training errors and norm of weight vectors. If the value of  $C$  selected is large enough, upper bound of the norm of weights is increased, and this leads to better classification. Indeed, regularisation of parameter  $C$  is able to control the influence of each individual support vector and prevent high overfitting in the recognition system. Overfitting occurs when the classifier model leads to low training errors and high generalisation errors in the data set. On the other hand, the width of the kernel surface is inversely proportional to parameter  $\gamma$ . When the value of  $\gamma$  is very small, a boundary for kernel RBF is larger than the maximum pairwise distance of data points, so SVM is not able to determine a certain decision boundary as overfitting is high. However, a large value of  $\gamma$  makes a simpler model with a low variance and high bias which does not tend to create an overfitting. The result tables show the effect of increasing the  $\gamma$  values on the test recognition rates causes substan-

**Table 3**A performance comparison of Gaussian OAA SVM classifiers with parameter value  $\gamma = 1$ .

DR	#F	False nonmatch rate		False match rate		Window recognition rate		Subject recognition rate	
		Train	Test	Train	Test	Train	Test	Train	Test
LDA	51	0 ± 0.0	95.5 ± 0.41	0 ± 0.0	0 ± 0.0	100 ± 0.0	4.48 ± 0.41	100 ± 0.0	0 ± 0.0
PCA	31	<b>0 ± 0.0</b>	<b>76.69 ± 1.37</b>	<b>0 ± 0.0</b>	<b>0.13 ± 0.13</b>	<b>100 ± 0.0</b>	<b>23.16 ± 1.3</b>	<b>100 ± 0.0</b>	<b>0 ± 0.0</b>
KPCA	54	0 ± 0.0	96.15 ± 0.55	0 ± 0.0	0 ± 0.0	100 ± 0.0	3.84 ± 0.55	100 ± 0.0	0 ± 0.0
	107	0 ± 0.0	99.58 ± 0.13	0 ± 0.0	0 ± 0.0	100 ± 0.0	0.4 ± 0.13	100 ± 0.0	0 ± 0.0
	111	0 ± 0.0	99.08 ± 0.34	0 ± 0.0	0 ± 0.0	100 ± 0.0	0.91 ± 0.34	100 ± 0.0	0 ± 0.0

**Table 4**A performance comparison of Gaussian OAA SVM classifiers with parameter value  $\gamma = 5$ .

DR	#F	C	False nonmatch rate		False match rate		Window recognition rate		Subject recognition rate	
			Train	Test	Train	Test	Train	Test	Train	Test
LDA	51	10	0 ± 0.0	21.42 ± 0.27	0 ± 0.0	1.18 ± 0.31	100 ± 0.0	77.37 ± 0.57	100 ± 0.0	22.43 ± 2.93
		60	0 ± 0.0	21.6 ± 0.21	0 ± 0.0	1.18 ± 0.31	100 ± 0.0	77.19 ± 0.41	100 ± 0.0	22.43 ± 2.93
		100	0 ± 0.0	21.6 ± 0.21	0 ± 0.0	1.18 ± 0.31	100 ± 0.0	77.19 ± 0.41	100 ± 0.0	22.43 ± 2.93
PCA	31	<b>10</b>	<b>0 ± 0.0</b>	<b>10.57 ± 1.12</b>	<b>0.15 ± 0.13</b>	<b>2.56 ± 0.78</b>	<b>99.84 ± 0.14</b>	<b>86.85 ± 1.63</b>	<b>97.43 ± 2.22</b>	<b>36.53 ± 5.09</b>
		60	0 ± 0.0	11.39 ± 1.19	0 ± 0.0	2.01 ± 0.48	100 ± 0.0	86.58 ± 1.62	100 ± 0.0	35.89 ± 4.45
		100	0 ± 0.0	11.49 ± 1.03	0 ± 0.0	1.96 ± 0.55	100 ± 0.0	86.53 ± 1.5	100 ± 0.0	35.89 ± 4.44
KPCA	54	10	0 ± 0.0	17.21 ± 0.28	0 ± 0.0	1.00 ± 0.07	100 ± 0.0	81.77 ± 0.2	100 ± 0.0	26.92 ± 1.92
		60	0 ± 0.0	17.3 ± 0.36	0 ± 0.0	0.95 ± 0.13	100 ± 0.0	81.72 ± 0.27	100 ± 0.0	26.28 ± 1.1
		100	0 ± 0.0	17.3 ± 0.36	0 ± 0.0	0.95 ± 0.13	100 ± 0.0	81.72 ± 0.27	100 ± 0.0	26.28 ± 1.1
	107	10	0 ± 0.0	35.2 ± 2.27	0 ± 0.0	0.59 ± 0.28	100 ± 0.0	64.19 ± 2.5	100 ± 0.0	12.81 ± 2.22
		60	0 ± 0.0	35.2 ± 2.27	0 ± 0.0	0.59 ± 0.28	100 ± 0.0	64.19 ± 2.5	100 ± 0.0	12.81 ± 2.22
		100	0 ± 0.0	35.2 ± 2.27	0 ± 0.0	0.59 ± 0.28	100 ± 0.0	64.19 ± 2.5	100 ± 0.0	12.81 ± 2.22
	111	10	0 ± 0.0	25.03 ± 1.37	0.12 ± 0.13	3.29 ± 0.47	99.87 ± 0.14	71.65 ± 1.59	98.07 ± 1.92	12.81 ± 2.93
		60	0 ± 0.0	25.08 ± 1.43	0.03 ± 0.05	3.24 ± 0.51	99.96 ± 0.05	71.65 ± 1.59	99.35 ± 1.11	12.81 ± 2.93
		100	0 ± 0.0	25.03 ± 1.37	0 ± 0.0	3.24 ± 0.51	100 ± 0.0	71.69 ± 1.55	100 ± 0.0	12.81 ± 2.93

**Table 5**A performance comparison of Gaussian OAA SVM classifiers with parameter value  $\gamma = 10$ .

DR	#F	C	False nonmatch rate		False match rate		Window recognition rate		Subject recognition rate	
			Train	Test	Train	Test	Train	Test	Train	Test
LDA	51	10	0 ± 0.0	10.43 ± 0.27	2.22 ± 0.61	6.04 ± 0.63	97.76 ± 0.61	83.51 ± 0.9	83.96 ± 2.93	35.89 ± 1.1
		60	0 ± 0.0	11.35 ± 0.7	0.03 ± 0.05	3.43 ± 1.1	99.96 ± 0.05	85.2 ± 1.46	99.35 ± 1.11	36.53 ± 3.84
		100	0 ± 0.0	11.34 ± 0.84	0 ± 0.0	3.47 ± 1.11	100 ± 0.0	85.16 ± 1.66	100 ± 0.0	36.53 ± 3.84
PCA	31	10	0 ± 0.0	5.21 ± 0.59	8.11 ± 0.56	12.03 ± 1.71	91.87 ± 0.56	82.73 ± 1.59	57.04 ± 4.00	40.38 ± 6.93
		60	0 ± 0.0	7.41 ± 0.55	0.54 ± 0.0	4.89 ± 1.38	99.45 ± 0.0	87.68 ± 1.76	92.94 ± 1.11	40.38 ± 3.85
		100	0 ± 0.0	7.73 ± 0.28	0.18 ± 0.18	4.48 ± 1.06	99.81 ± 0.18	87.77 ± 0.95	96.79 ± 2.93	40.38 ± 0.0
KPCA	54	10	0 ± 0.0	8.64 ± 0.47	1.55 ± 0.47	4.02 ± 0.65	98.43 ± 0.47	87.31 ± 0.48	86.53 ± 3.33	36.53 ± 5.09
		<b>60</b>	<b>0 ± 0.0</b>	<b>9.24 ± 0.31</b>	<b>0.03 ± 0.05</b>	<b>2.56 ± 0.57</b>	<b>99.96 ± 0.5</b>	<b>88.18 ± 0.82</b>	<b>99.35 ± 1.11</b>	<b>40.38 ± 5.08</b>
		100	0 ± 0.0	9.33 ± 0.23	0 ± 0.0	2.56 ± 0.67	100 ± 0.0	88.09 ± 0.84	100 ± 0.0	40.38 ± 5.08
	107	10	0 ± 0.0	13.04 ± 0.68	0.03 ± 0.05	1.18 ± 0.16	99.96 ± 0.05	85.75 ± 0.82	99.35 ± 1.11	32.69 ± 0.0
		60	0 ± 0.0	13.32 ± 0.83	0 ± 0.0	1.09 ± 0.36	100 ± 0.0	85.57 ± 1.09	100 ± 0.0	30.12 ± 2.22
		100	0 ± 0.0	13.32 ± 0.83	0 ± 0.0	1.09 ± 0.36	100 ± 0.0	85.57 ± 1.09	100 ± 0.0	30.12 ± 2.22
	111	10	0 ± 0.0	10.02 ± 1.48	0.33 ± 0.18	6.45 ± 0.96	99.66 ± 0.18	83.51 ± 1.43	94.22 ± 1.92	29.48 ± 2.93
		60	0 ± 0.0	10.15 ± 1.62	0.09 ± 0.09	6.08 ± 0.87	99.9 ± 0.09	83.74 ± 1.66	98.07 ± 1.92	30.12 ± 4.00
		100	0 ± 0.0	10.15 ± 1.52	0.06 ± 0.1	6.04 ± 0.83	99.9 ± 0.1	83.78 ± 1.68	98.71 ± 2.22	30.12 ± 4.00

tial lower overfitting through decreasing False Nonmatch Rate in spite of a rise in False Match Rate. Fig. 6 shows the effect on  $(C, \gamma)$  in window predictive rates (for test data) for all the data sets.

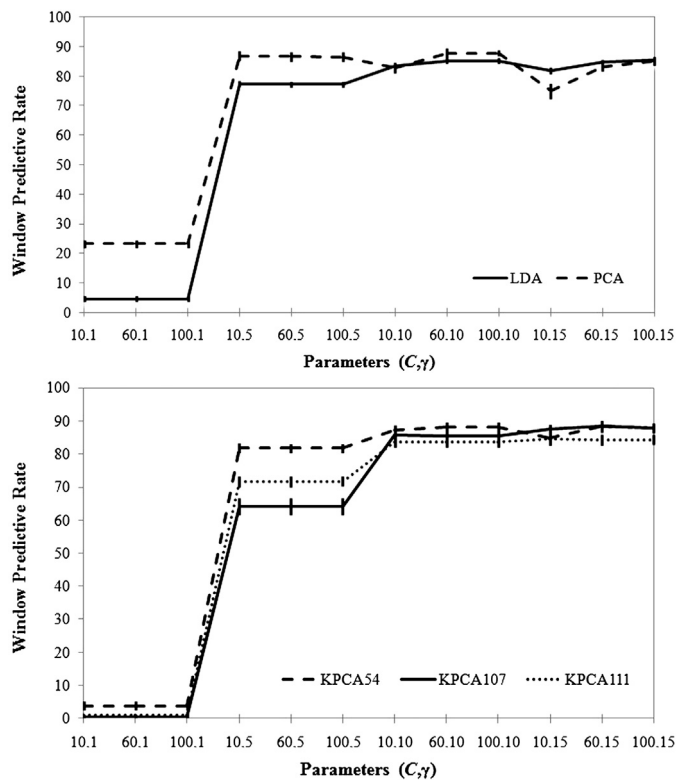
It can be seen from Table 3 that the Gaussian OAA SVM model cannot be robust for biometric recognition because the highest generalisation error in the test data sets occurs when unsuitable value of the kernel parameter is chosen. In other words, the classifiers may well learn training data with the lowest biometric errors, but the worst recognition rates occur in all testing data. The different values of parameter  $C$  are not effective for recognition performance system and in attaining similar accuracy, so its values are disregarded in Table 3. The value highlighted in bold specifies the best result obtained in the different tables.

An effect that can be examined is whether the relationship between the optimal  $\gamma$  value of RBF kernel classification corresponds to the best recognition rates and the number of features in the

data sets. Although the values of  $\gamma$  parameter and the number of features are not directly related, the best possible  $\gamma$  values mainly occur in data sets with a small number of features. The Gaussian kernel value computed with  $\exp(-\gamma \|x_i - x_j\|^2)$  is close to a zero for data sets with a large number of features (attributes) and large  $\gamma$  values. Hence, it cannot adequately generalise a large number of support vectors. Due to this fact, a direct correlation between increasing the number of features and the  $\gamma$  value up to 15 and the recognition performance of the KPCA data sets as in Tables 3, 4, 5 and 6 can be observed. However, the different variations captured (related to eigenvalues) by using KPCA for extraction of significant features can also be effective in determining the accuracy of the Gaussian OAA SVM models. The verification recognition rates (WRR and SRR) are illustrated by the  $KPCA111 < KPCA107 < KPCA54$  as the  $\gamma$  value and the number of attributes is increased. In addition, the result shows that an optimal  $\gamma$  value of KPCA data sets is 15.

**Table 6**A performance comparison of Gaussian OAA SVM classifiers with parameter value  $\gamma = 15$ .

DR	#F	C	False nonmatch rate		False match rate		Window recognition rate		Subject recognition rate	
			Train	Test	Train	Test	Train	Test	Train	Test
LDA	51	10	0 $\pm$ 0.0	6.91 $\pm$ 0.31	7.17 $\pm$ 0.68	11.26 $\pm$ 1.1	92.82 $\pm$ 0.68	81.81 $\pm$ 0.84	53.2 $\pm$ 2.21	35.25 $\pm$ 1.1
		60	0 $\pm$ 0.0	9.06 $\pm$ 0.41	1.3 $\pm$ 0.68	6.31 $\pm$ 0.13	98.68 $\pm$ 0.68	84.61 $\pm$ 0.36	86.53 $\pm$ 1.92	37.81 $\pm$ 4.00
		100	0 $\pm$ 0.0	9.38 $\pm$ 0.2	0.21 $\pm$ 0.18	5.26 $\pm$ 0.57	99.78 $\pm$ 0.19	85.34 $\pm$ 0.48	96.15 $\pm$ 3.33	37.17 $\pm$ 4.00
PCA	31	10	0 $\pm$ 0.0	2.78 $\pm$ 0.55	18.73 $\pm$ 0.52	22.29 $\pm$ 2.59	81.25 $\pm$ 0.52	74.9 $\pm$ 2.55	35.89 $\pm$ 1.1	32.04 $\pm$ 2.93
		60	0 $\pm$ 0.0	5.99 $\pm$ 0.48	5.61 $\pm$ 0.36	9.68 $\pm$ 0.09	94.37 $\pm$ 0.36	83.19 $\pm$ 1.53	66.66 $\pm$ 1.1	37.17 $\pm$ 2.22
		100	0 $\pm$ 0.0	6.45 $\pm$ 0.27	3.04 $\pm$ 0.45	8.6 $\pm$ 1.5	96.94 $\pm$ 0.45	84.93 $\pm$ 1.46	76.92 $\pm$ 5.77	40.38 $\pm$ 1.92
KPCA	54	10	0 $\pm$ 0.0	6 $\pm$ 0.56	5.36 $\pm$ 0.36	9.06 $\pm$ 0.76	94.62 $\pm$ 0.36	84.93 $\pm$ 0.75	65.39 $\pm$ 3.3	39.74 $\pm$ 4.84
		60	0 $\pm$ 0.0	<b>7.59 <math>\pm</math> 0.2</b>	<b>0.36 <math>\pm</math> 0.24</b>	<b>3.97 <math>\pm</math> 0.36</b>	<b>99.62 <math>\pm</math> 0.24</b>	<b>88.41 <math>\pm</math> 0.28</b>	<b>92.94 <math>\pm</math> 4.00</b>	<b>41.66 <math>\pm</math> 4.00</b>
		100	0 $\pm$ 0.0	8.19 $\pm$ 0.15	0.15 $\pm$ 0.05	3.88 $\pm$ 0.41	99.84 $\pm$ 0.05	87.9 $\pm$ 0.47	96.79 $\pm$ 1.1	41.66 $\pm$ 4.84
	107	10	0 $\pm$ 0.0	9.15 $\pm$ 0.93	0.79 $\pm$ 0.55	3.24 $\pm$ 0.31	99.2 $\pm$ 0.55	87.58 $\pm$ 1.24	92.94 $\pm$ 4.83	39.1 $\pm$ 1.1
		60	0 $\pm$ 0.0	9.98 $\pm$ 0.68	0 $\pm$ 0.0	1.68 $\pm$ 1.03	100 $\pm$ 0.0	88.32 $\pm$ 1.68	100 $\pm$ 0.0	40.38 $\pm$ 0.0
		100	0 $\pm$ 0.0	8.19 $\pm$ 0.15	0.15 $\pm$ 0.05	3.88 $\pm$ 0.41	99.84 $\pm$ 0.05	87.9 $\pm$ 0.47	96.79 $\pm$ 1.1	41.66 $\pm$ 4.84
	111	10	0 $\pm$ 0.0	7.13 $\pm$ 1.78	1.73 $\pm$ 0.24	8.33 $\pm$ 0.55	98.25 $\pm$ 0.24	84.51 $\pm$ 1.86	76.92 $\pm$ 0.0	28.84 $\pm$ 5.08
		60	0 $\pm$ 0.0	8.92 $\pm$ 1.48	0.15 $\pm$ 0.1	6.82 $\pm$ 0.57	99.84 $\pm$ 0.1	84.24 $\pm$ 1.59	97.43 $\pm$ 1.1	25.63 $\pm$ 4.00
		100	0 $\pm$ 0.0	9.06 $\pm$ 1.48	0.15 $\pm$ 0.1	6.63 $\pm$ 0.41	99.84 $\pm$ 0.1	84.29 $\pm$ 1.5	97.43 $\pm$ 1.1	25.63 $\pm$ 4.00

**Fig. 6.** The effect of parameters ( $C, \gamma$ ) of Gaussian OAA SVMs in window recognition rate over test data sets: linear (up-side) and nonlinear (down-side) dimension reduction methods.

However, these tables do not report the recognition rates which decline.

Moreover, the relationship between the optimal  $\gamma$  value and the best recognition performance of both LDA and PCA features is similar to what have been explained earlier. The recognition rates increase when the  $\gamma$  values go up to 15 and 10 for LDA and PCA features respectively, after which they slowly decline. Nevertheless, a general correlation cannot be found between the optimal RBF  $\gamma$  and the number of features and the best recognitions on the experiment results of all the data sets. The highest window and subject predictive accuracies are achieved to approximately 88.41% and 41.66% by using the Gaussian OAA SVM classifier which is set at the optimal setting value  $(C, \gamma) = (60, 15)$  on the KPCA data set with ( $\#F = 54$ ) that has the lowest average biometric errors and

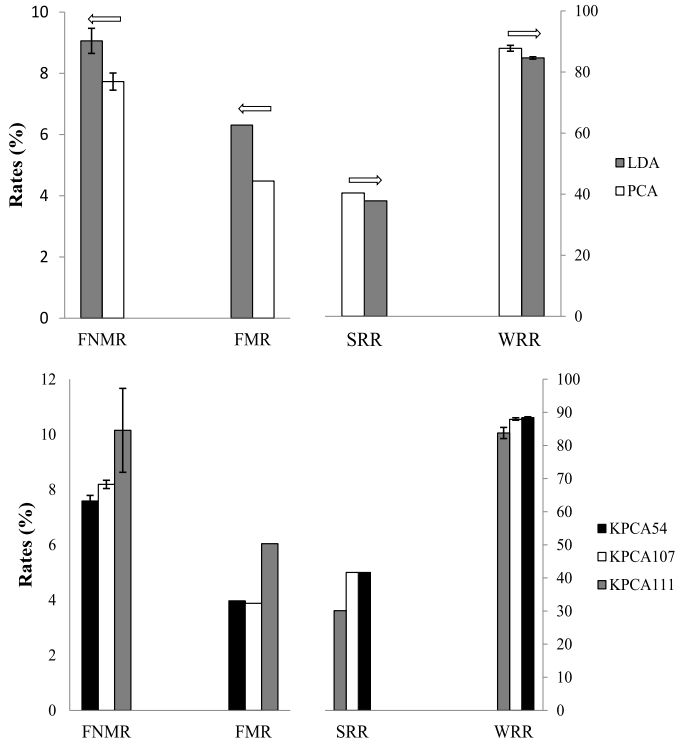
less overfitting than others. The biometric errors of FNMR and FMR are decreased to about 7.59 and 3.97%, respectively on the KPCA54 data set.

The low subject recognition rate seems to have high unclassified samples which do not fit into any model. The highly skewed the data sets affect to perform classifiers due to poor bias approximation. Therefore, a brief investigation can also be carried out on the influence of FNMR and FMR errors on the biometric recognition rates of WRR and SRR in Tables 3, 4, 5 and 6. FNMR errors are reduced for each of the data sets with an increase in the  $\gamma$  values, so the window recognition rates (WRRs) as well as subject recognition rates (SRRs) are increased until the optimal  $\gamma$  value for each data is set. It can be seen from the tables that the window recognition rate may be reduced in some data sets, but subject recognition rate can be raised or vice-versa.

Fig. 7 clearly shows the effect of biometric errors for the best subject recognition rates on LDA and PCA and all KPCA test data sets.

Another effect that can be considered is whether the improvement in performance is due to the denoising approach or the application of KPCA. KPCA is more effective in increasing performance and good generalisation of the system compared with LDA and PCA methods according to above discussion. In order to evaluate the impact of denoising approach, the results of Gaussian OAA SVM classifier based on KPCA feature extraction on both the denoised individual ECG signals using *coif3* and *db7* wavelets are compared. According to Tables 1 and 2, the best performance of *db7* wavelet is at level 4 with Rigrsure rule of hard thresholding. The experimental results given in Table 7 indicate that the effect of different mother wavelets on error and recognition rates of Gaussian OAA SVMs with  $(C, \gamma) = (60, 15)$  on test data including 54 features. The optimal selection of mother wavelet can be crucial to have signals with the lowest reconstruction errors and highest Signal-to-Noise Ratio for achieving higher recognition rates in the ECG based biometrics.

A comparison has been made between computational complexity and memory requirements which involve dimensionality reduction algorithms. The elapsed time needed to perform each reducing dimension processing of the techniques on the high dimensional space of AC windows (samples) has also been evaluated. In addition, the average elapsed training time as well as testing time needed to build the Gaussian OAA SVM classifiers for all subjects on the ECG data sets with significant dimensions and predict test samples on the SVM models have also been considered, and these are shown in Table 8.



**Fig. 7.** The effect of biometric errors on the best subject recognition rates of all test data sets: linear (up-side) and nonlinear (down-side) dimension reduction methods.

**Table 7**

A predictive performance comparison of Gaussian OAA SVM on denoised ECG data sets using discrete wavelet transforms.

Error & recognition rates	Mother wavelet	
	coif3	db7
False nonmatch rate	<b>7.59 ± 0.20</b>	9.06 ± 0.71
False match rate	<b>3.97 ± 0.36</b>	5.67 ± 0.55
Window recognition rate	<b>88.41 ± 0.28</b>	85.25 ± 0.15
Subject recognition rate	<b>41.66 ± 4.00</b>	32.68 ± 3.33

The computational complexity and memory required for the dimension reduction techniques LDA, PCA and KPCA are determined by two data properties – the number of data samples  $n$  and the original dimension space  $d$ . LDA involves the eigenvalue problem that can be especially expensive for both computational and memory. LDA has  $O(ndt + t^3)$  computational and requires  $O(nd + nt + dt)$  memory, where  $t = \min(n, d)$  [51]. The computational complexity of PCA to solve eigenvalue problem of the  $d \times d$  covariance matrix is  $O(d^3)$ . The memory required of PCA is  $O(d^2)$  to store the  $d \times d$  covariance matrix. The nonlinear method of KPCA has

**Table 8**

Average elapsed time needed to build Gaussian SVMs on ECG training data and recognition ECG test data (in seconds).

DR	C	$\gamma = 1$	$\gamma = 5$	$\gamma = 10$	$\gamma = 15$
Training time					
LDA	10	0.6	0.21	0.12	0.12
	60	0.6	0.21	0.12	0.12
	100	0.6	0.21	0.12	0.12
PCA	10	0.52	0.11	0.12	0.11
	60	0.52	0.11	0.14	0.14
	100	0.54	0.11	0.13	0.15
KPCA54	10	0.57	0.12	0.11	0.11
	60	0.56	0.12	0.12	0.12
	100	0.57	0.12	0.11	0.13
KPCA107	10	0.6	0.21	0.12	0.11
	60	0.59	0.21	0.12	0.12
	100	0.59	0.21	0.12	0.12
KPCA111	10	0.59	0.24	0.11	0.11
	60	0.59	0.25	0.11	0.12
	100	0.59	0.25	0.12	0.13
Testing time					
LDA		0.07	0.04	0.03	0.03
PCA		0.05	0.02	0.01	0.01
KPCA54		0.06	0.03	0.02	0.02
KPCA107		0.07	0.04	0.03	0.03
KPCA111		0.07	0.04	0.03	0.03

computational and memory complexity disadvantages compared to PCA method. The KPCA method performs an eigenvalue problem with the  $n \times n$  kernel matrix and demands for a memory including square with the number of data samples  $n$ . Therefore, the computational and memory complexity of KPCA are  $O(n^3)$  and  $O(n^2)$ , respectively [26]. In this paper, most of elapsed time, in seconds, for the linear and nonlinear dimension reduction techniques in which  $KPCA111 > KPCA107 > KPCA54 > LDA > PCA$  with time values of 16.18, 15.16, 5.45, 0.2, 0.05 s, respectively. It can be seen from Table 8 that the optimal  $\gamma$  value for each of the data set is able to reduce complexity, and so the average elapsed tainting time to learn the Gaussian SVM classifiers for all subjects is decreased. There is no significant difference in the time of data sets in the optimal setting parameters ( $C, \gamma$ ). In addition, recognition of test data is faster as shown in Table 8. The value of the  $C$  parameter does not affect the testing computational time of the SVM models, so the values have been disregarded in the table.

In order to evaluate the proposed framework validation, a performance comparison between random partition holdout (60–40%) and 10-fold cross-validation techniques is summarised in Table 9 for both one and two recording sessions. With regard to Section 2.3.1, the total windows for autocorrelation of the 32 ECG signals (two records for 16 subjects) are 1120 with the vector

**Table 9**

A predictive performance comparison of Gaussian OAA SVM with random partitioning methods on one and two recording sessions.

DR	#F	C	$\gamma$	#Session	Random partitioning method							
					10-Fold cross-validation				Holdout (60–40% split)			
					FNMR	FMR	WRR	SRR	FNMR	FMR	WRR	SRR
LDA	51	60	15	One <sup>1</sup>	6.22 ± 1.75	4.59 ± 1.62	89.15 ± 1.88	71.79 ± 4.44	9.06 ± 0.41	6.31 ± 0.13	84.61 ± 0.36	37.81 ± 4.00
	15	25	10	Two <sup>2</sup>	1.12 ± 1.07	4.57 ± 1.63	94.28 ± 1.85	71.45 ± 4.45	1.11 ± 0.0	6.69 ± 1.17	92.18 ± 1.18	27.08 ± 9.54
PCA	31	100	10	One	6.22 ± 1.81	3.47 ± 1.51	90.28 ± 2.8	74.09 ± 6.36	7.73 ± 0.28	4.48 ± 1.06	87.77 ± 0.95	40.38 ± 0.0
	26	25	8	Two	3.5 ± 1.68	3.23 ± 1.43	93.05 ± 2.06	65.2 ± 11.4	3.42 ± 1.3	3.84 ± 0.53	92.7 ± 1.81	20.83 ± 13.01
KPCA	54	60	15	One	<b>4.83 ± 1.57</b>	<b>3.52 ± 1.57</b>	<b>91.62 ± 2.05</b>	<b>76.34 ± 5.27</b>	<b>7.59 ± 0.2</b>	<b>3.97 ± 0.36</b>	<b>88.41 ± 0.28</b>	<b>41.66 ± 4.00</b>
				Two	<b>2.97 ± 1.69</b>	<b>2.46 ± 1.42</b>	<b>94.54 ± 2.31</b>	<b>71.66 ± 10.85</b>	<b>3.34 ± 1.16</b>	<b>2.52 ± 0.25</b>	<b>94.12 ± 1.05</b>	<b>29.16 ± 13.01</b>

<sup>1</sup> One recording session for 52 subjects.

<sup>2</sup> Two recording sessions for 16 subjects.



**Table 10**  
Comparison of the current framework and other studies.

Reference	ECG database	#P	SL, SS (min, sec)	Condition	#S	FT	Preprocessing	FE	Classification	Error (%)	Accuracy (%)
[4]	Non-clinical	16	3, 5	Rest	Two	Non-fiducial	Butterworth bandpass filter	AC/LDA	NN on Euclidean distance	EER: 14	–
[28]	Non-clinical	52	3, 5	Rest	One	Non-fiducial	Butterworth bandpass filter	AC/LDA, PT <sup>†</sup>	NN on Euclidean distance	–	92.3
[34]	Non-clinical	32	5, 1–5	Rest	One	Hybrid	Finite impulse response filter	Adaptive threshold	KNN	2 < EER < 10	–
									SVM	13 < FNMR < 51 FMR: 0	–
[35]	Non-clinical	26	5, 25	Exercise Rest	One	Non-fiducial	Chaos theory		SVM-Exercise SVM-Rest SVM-E&R	–	81.73 71.79 63.62
[6]	Clinical	56	2905*, 10	Healthy	One	Non-fiducial	Butterworth bandpass filter	AC/LDA	NN on Euclidean distance	–	WRR: 96.2 SRR: 96.42
Proposed method	Non-clinical	52	3, ~ 5	Rest	One	Non-fiducial	Discrete wavelet transform	AC/KPCA	SVM-holdout SVM-cross validation SVM-holdout SVM-cross validation	FNMR: 7.59 4.83 3.34 2.97 FMR: 3.97 3.5 2.52 2.46 SRR: 88.41 91.62 94.12 94.54	41.66 76.34 29.16 71.66

P: Person; SL: Signal Length; SS: Signal Segment; S: Session; FT: Feature Type; FE: Feature Extraction.

<sup>†</sup> PT: Periodicity Transform.

\* Number of windows.

of dimension 160 for every window. In this experiment, the dimensionality of the generic data set is respectively reduced from 160 to 15, 26 and 54 using LDA, PCA and KPCA methods. PCA and Gaussian based KPCA in  $\gamma$  value of 0.01 which retain 99% of the data's variance. According to holdout method (60–40%), each data set is randomly split into the 672 training data and remaining 448 are test sets for Gaussian OAA SVM classification. However, 10-fold cross-validation is used to partition randomly each of the ECG data sets predefined into 10 equal-sized multi-class subsets. This process is repeated 10 times so that every partition is utilised for testing exactly once. Although this approach leads to estimate accurate performance, it is computationally expensive to repeat the procedure 10 times. As Table 9 shown, 10-fold cross-validation is an effective random partition method for predictive recognition rates, because the technique can reduce the variance [52]. The highest window and subject predictive accuracies are achieved on the KPCA data sets for each one and two recording sessions in both random partition methods that have the lowest average biometric errors and less overfitting than others. On the other hand, the lowest rates of subject recognition for both random partition methods on two recording sessions could be the result of variations within ECG waveforms over time-varying. Based on Table 9, when the SVM classifier deals with a large number of classes, window recognition rates are decreased due to poor bias approximation.

A comparison of the current framework with some methodologies found in literature is summarised in Table 10. Although the two main factors of clinical source data and length of ECG segment were highly effective in increasing accuracy of identity recognition in study [6], it is impractical for real life applications in condition where recognition must be processed quickly.

With regard to Table 10, the number of subjects for the SVM based studies was small compared with other works [34,35]. Lourenço et al. has shown that SVM can be a powerful method for ECG-based recognition in small number of subjects [34]. The current methodology outperforms for a large number of single-session in resting status while the recognition performance is mostly degraded by increasing number of subjects. It seems that discarding of the non-significant features such as noise and redundant features can affect the SVM recognition performance which includes classification accuracy and computational efficiency.

As experimental results shown in Table 10, the performance of current framework is comparable with study [28] for 52 subjects. The study [28] was demonstrated that the periodicity transform can be considerably useful for ECG analysis in order to outlier detection. On the other hand, the proposed method and the work [4] were only tested over 16 subjects in two recording sessions, which can be more considered in multiple sessions with large number of recordings for different postures and physiological conditions in future.

## 5. Conclusion

We proposed a new formulation for biometric authentication system based on the one-lead electrocardiogram signal using the kernel methods. Denoising on the noisy ECG signals (non-clinical source data) was performed initially by Discrete Wavelet Transform with *coif3* at level 4, followed by an evaluation of feature extraction and classification. This study is based on non-fiducial approach to extract the ECG features of a population where there is a large number of redundant information. The experiments of verification recognition system have been tested over random unknown ECG data sets designed in biometric feature extraction phase. Autocorrelation method was applied to extract discriminative patterns with high dimensionality from the ECG windows according to the previous methodologies. A systematic analysis shows that

the autocorrelation of ECGs and a Gaussian KPCA method as feature extractor has a maximum effect on achieving high subject and window recognition rates of Gaussian OAA SVMs compared to linear dimension reduction techniques (PCA, LDA). However, the computation time of SVM recognition models is almost identical for all dimension reduction techniques. The proposed method is feasible and offers a secure solution to ECG verification system to control a trade-off between False Match and False Nonmatch rates when dealing with highly unknown subjects. Further experiments are needed to optimise the individual recognition system based on kernel SVM methods and make it more robust and further improves classifier generalisation ability in ECG based biometric systems.

## Acknowledgments

This research has been supported by the Ministry of Education Malaysia (MOE) through Fundamental Research Grant Scheme (FRGS) under project code 5524616. The authors would like to thank the University of Toronto to make an agreement with Universiti Putra Malaysia (UPM) for applying ECG database collected at the BioSec.Lab.

## References

- [1] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, *IEEE Trans. Circuits Syst. Video Technol.* 14 (1) (2004) 4–20.
- [2] Y. Wang, F. Agraftioti, D. Hatzinakos, Analysis of human electrocardiogram for biometric recognition, *EURASIP J. Adv. Signal Process.* 2008 (2008) 148658.
- [3] L. Biel, O. Pettersson, L. Philipson, P. Wide, ECG analysis: a new approach in human identification, *IEEE Trans. Instrum. Meas.* 50 (3) (2001) 808–812.
- [4] F. Agraftioti, J. Gao, D. Hatzinakos, Heart biometrics: theory, methods and application, in: J. Yang (Ed.), *Biometrics*, 1st ed., InTech, 2011, pp. 199–216.
- [5] Y. Wang, K.N. Plataniotis, D. Hatzinakos, Integrating analytic and appearance attributes for human identification from ECG signal, in: *Proc. Biometrics Symposiums*, Baltimore, USA, September 2006, pp. 1–6.
- [6] F. Agraftioti, D. Hatzinakos, ECG biometric analysis in cardiac irregularity conditions, *Signal Image Video Process.* 3 (4) (2008) 329–343.
- [7] F. Agraftioti, F.M. Bui, D. Hatzinakos, Medical biometrics in mobile health monitoring, *J. Secur. Commun. Netw.* 4 (5) (July 2010) 525–539.
- [8] R.D. Labati, V. Piuri, R. Sassi, F. Scotti, G. Sforza, Adaptive ECG biometric recognition: a study on re-enrollment methods for QRS signals, in: *Proc. IEEE CIBIM*, Orlando, FL, USA, 2014, pp. 30–37.
- [9] M.A. Haque, K. Nasrollahi, T.B. Moeslund, Can contact-free measurement of heartbeat signal be used in forensics?, in: *Proc. IEEE EUSIPCO*, Nice, France, 2015, pp. 769–773.
- [10] J. Gao, F. Agraftioti, H. Mohammadzade, D. Hatzinakos, ECG for blind identity verification in distributed systems, in: *Proc. IEEE ICASSP*, Prague, Czech, 2011, pp. 1916–1919.
- [11] I. Odinaka, P. Lai, A.D. Kaplan, J.A. O'Sullivan, E.J. Sirevaag, J.W. Rohrbaugh, ECG biometric recognition: a comparative analysis, *IEEE Trans. Inf. Forensics Secur.* 7 (6) (November 2012) 1812–1824.
- [12] F. Sufi, I. Khalil, J. Hu, ECG-based authentication, in: P. Stavroulakis, M. Stamp (Eds.), *Handbook of Information and Communication Security*, Springer, Berlin, Heidelberg, 1sted., 2010, pp. 309–331.
- [13] F. Odone, M. Pontil, A. Verri, Machine learning techniques for biometrics, in: M. Tistarelli, S.Z. Li, R. Chellappa (Eds.), *Handbook of Remote Biometrics*, Advances in Pattern Recognition, 1st ed., Springer, London, 2009, pp. 247–271.
- [14] S.A. Israel, J.M. Irvine, A. Cheng, M.D. Wiederhold, B.K. Wiederhold, ECG to identify individuals, *Pattern Recognit.* 38 (1) (August 2005) 133–142.
- [15] Y.N. Singh, P. Gupta, ECG to individual identification, in: *Proc. 2nd IEEE BTAS 2008*, Arlington, USA, 2008, pp. 1–8.
- [16] M.S. Islam, N. Alajlan, Y. Bazi, H.S. Hichri, HBS: a novel biometric feature based on heartbeat morphology, *IEEE Trans. Inf. Technol. Biomed.* 16 (3) (May 2012) 445–453.
- [17] D.P. Coutinho, H. Silva, H. Gamboa, A. Fred, M. Figueiredo, Novel fiducial and non-fiducial approaches to electrocardiogram-based biometric systems, *IET Biometrics* 2 (2) (February 2013) 64–75.
- [18] C. Carreiras, A. Lourenço, H. Silva, A. Fred, A unifying approach to ECG biometric recognition using the wavelet transform, in: M. Kamel, A. Campilho (Eds.), *Image Analysis and Recognition*, Springer, London, 2013, pp. 53–62.
- [19] A. Lourenço, H. Silva, A. Fred, Unveiling the biometric potential of finger-based ECG signals, *Comput. Intell. Neurosci.* 2011 (2011) 720971, 1–8.
- [20] H. Silva, A. Lourenço, F. Canento, A.L. Fred, N. Raposo, ECG biometrics: principles and applications, in: *Proc. BIOSIGNALS*, Lisbon, Portugal, 2013, pp. 215–220.
- [21] H.P. da Silva, A. Lourenço, A. Fred, N. Raposo, M. Aires-de-Sousa, Check your biosignals here: a new dataset for off-the-person ECG biometrics, *Comput. Methods Programs Biomed.* 113 (2) (February 2014) 503–514.
- [22] H.P. da Silva, C. Carreiras, A. Lourenço, A. Fred, R.C. das Neves, R. Ferreira, Off-the-person electrocardiography: performance assessment and clinical correlation, *Health Technol.* 4 (4) (April 2015) 309–318.
- [23] K.N. Plataniotis, D. Hatzinakos, J.K.M. Lee, ECG biometric recognition without fiducial detection, in: *Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, 2006, pp. 1–6.
- [24] A.D. Chan, M.M. Hamdy, A. Badre, V. Badee, Wavelet distance measure for person identification using electrocardiograms, *IEEE Trans. Instrum. Meas.* 57 (2) (February 2008) 248–253.
- [25] J. Yang, Z. Jin, J. Yang, Non-linear techniques for dimension reduction, in: A. Jain, S.Z. Li (Eds.), *Encyclopedia of Biometrics*, Springer US, 2009, pp. 1003–1007.
- [26] L.J.P. Van der Maaten, E.O. Postma, H.J. Van den Herik, Dimensionality reduction: a comparative review, Technical report, TiCC-TR 2009-005, Tilburg University, October 2009.
- [27] F. Agraftioti, D. Hatzinakos, Fusion of ECG sources for human identification, in: *Proc. IEEE ISCCSP*, March 2008, pp. 1542–1547.
- [28] F. Agraftioti, D. Hatzinakos, Signal validation for cardiac biometrics, in: *Proc. IEEE ICASSP*, Dallas, USA, March 2010, pp. 1734–1737.
- [29] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, K.R. Müller, Fisher discriminant analysis with kernels, in: *Neural Networks for Signal Processing IX*, Proc. IEEE Signal Processing Society Workshop, August 1999, pp. 41–48.
- [30] B. Schölkopf, A. Smola, *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, USA, 2002, pp. 405–468.
- [31] B. Schölkopf, A. Smola, K.R. Müller, Kernel principal component analysis, in: B. Schölkopf, C.J.C. Burges, A.J. Smola (Eds.), *Advances in Kernel Methods: Support Vector Learning*, 1st ed., MIT Press, 1999, pp. 327–352.
- [32] P.N. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*, 1st ed., Pearson Addison-Wesley, Boston, 2006, pp. 256–309.
- [33] J. Sriram, M. Shin, T. Choudhury, D. Kotz, Activity-aware ECG-based patient authentication for remote health monitoring, in: *Proc. ICMI-MLMI*, New York, USA, November 2009, pp. 297–304.
- [34] A. Lourenço, H. Silva, A. Fred, ECG-based biometrics: a real time classification approach, in: *IEEE International Workshop on Machine Learning for Signal Processing*, MLSP, Santander, Spain, September 2012, pp. 1–6.
- [35] S.L. Lin, C.K. Chen, C.L. Lin, W.C. Yang, C.T. Chiang, Individual identification based on chaotic electrocardiogram signals during muscular exercise, *IET Biometrics* 3 (4) (December 2014) 257–266.
- [36] Vapnik, S. Golowich, A. Smola, Support vector method for function approximation, regression estimation, and signal processing, in: M. Mozer, M. Jordan, T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9*, MIT Press, 1997, pp. 281–287.
- [37] E.B. Lin, M. Haske, M. Smith, D. Sowards, Wavelet analysis of ECG signals, in: X. Shen, A.I. Zayed (Eds.), *Multiscale Signal Analysis and Modeling*, 1st ed., Springer-Verlag, New York, 2013, pp. 233–255.
- [38] M. AlMahamdy, H.B. Riley, Performance study of different denoising methods for ECG signals, in: *Proc. ICTH-2014*, Athens 45701, USA, 2014, pp. 325–332.
- [39] B.N. Singh, A.K. Tiwari, Optimal selection of wavelet basis function applied to ECG signal denoising, *Digit. Signal Process.* 16 (3) (January 2006) 275–287.
- [40] M. Abo-Zahhad, ECG signal compression using discrete wavelet transform, in: J.T. Oikkonen (Ed.), *Discrete Wavelet Transforms – Theory and Applications*, 1st ed., InTech, 2011, pp. 143–168.
- [41] D.L. Donoho, De-noising by soft-thresholding, *IEEE Trans. Inf. Theory* 41 (3) (1995) 613–627.
- [42] D.L. Donoho, J.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, *Biometrika* 81 (3) (1992) 425–455.
- [43] L. Birgé, P. Massart, From model selection to adaptive estimation, in: D. Pollard, E. Torgersen, G.L. Yang (Eds.), *Festschrift for Lucien Le Cam Research Papers in Probability and Statistics*, 1st ed., Springer, New York, 1997, pp. 55–87.
- [44] J. Ye, S. Ji, Discriminant analysis for dimensionality reduction: an overview of recent developments, in: N.V. Boulgouris, K.N. Plataniotis, E. Micheli-Tzanakou (Eds.), *Biometrics: Theory, Methods, and Applications*, 1st ed., Wiley-IEEE Press, 2009, pp. 1–19.
- [45] K.R. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf, An introduction to kernel-based learning algorithms, *IEEE Trans. Neural Netw.* 12 (2) (Aug 2001) 181–201.
- [46] B. Schölkopf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.* 10 (1) (July 1998) 1299–1319.

- [47] T. Scheidat, M. Leich, M. Alexander, C. Vielhauer, Support vector machines for dynamic biometric handwriting classification, in: *Proc. CEUR Workshop, AIAI Workshops, Greece, 2009*, pp. 118–125.
- [48] V.N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed., Springer, New York, 2000, pp. 17–259.
- [49] S. Abe, *Support Vector Machines for Pattern Classification*, 2nd ed., Springer-Verlag, London, 2010, pp. 21–158.
- [50] P.H. Chen, C.J. Lin, B. Schölkopf, A tutorial on  $\nu$ -support vector machines, *Appl. Stoch. Models Bus. Ind.* 21 (2) (2005) 111–136.
- [51] D. Cai, X. He, J. Han, SRDA: an efficient algorithm for large-scale discriminant analysis, *IEEE Trans. Knowl. Data Eng.* 20 (1) (Jan 2008) 1–12.
- [52] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Proc. IJCAI*, vol. 2, Quebec, Canada, August 1995, pp. 1137–1145.



**Maryamsadat Hejazi** received B.Sc. degree in computer software engineering from University of Applied Science and Technology, Iran, in 2008, and graduated M.Sc. in software engineering and software architecture from Multimedia University, Malaysia, in 2012. She is currently working toward Ph.D. degree in biometric systems at Engineering faculty, Universiti Putra Malaysia. Her research interests are intelligent systems, machine learning and data mining and their applications to classification, anomaly detection and biometric systems.



**S.A.R. Al-Haddad** is a Ph.D. graduated in Electrical, Electronic and Systems Engineering from National University Malaysia. His specialised in Human and Animal sound Processing, Al-Quran Sound Processing, Media Security and Biometric. He is a lecturer in Department of Computer and Communications Systems Engineering, Universiti Putra Malaysia since 1997 and promoted as Associate Professor in year 2012. He taught students for undergraduate and graduate and managed to get International and national grants. Further than that, he has few patents and copyrights and actively joins professional society such as Malaysia IEEE as Senior Member and National Islamic Calligraphy as Chairman of IT.



**Yashwant Prasad Singh** received the B.Sc. Eng. (Electrical) degree from Bihar College of Engineering, India, the M.E. degree in electronics and communications (first-class honours) from the University of Roorkee, India, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology, India, in 1966, 1975, and 1980, respectively. He has worked in the field of computer science for over 45 years and has published over 150 technical papers in international and national journals and conference proceedings. He has held various academic positions in various Indian and Malaysian Universities and had been professor of computer science, Faculty of Computing and Informatics, Multimedia University, Malaysia during May, 1998 till February 2014. He had also held Professor/Adjunct Professor in MRU, Faridabad (India) during February 2014–January 2016. He now works as an independent researcher in the areas of Artificial Intelligence algorithms and Machine learning and their applications to classification, clustering, anomaly detection, association rule mining, ranking and routing, big data analytics.



**Shaiful Jahari Hashim** is currently an Associate Professor in the Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia. He received his Ph.D. from Cardiff University, UK (2011), M.Sc. from National University of Malaysia (2003) and B.Eng. from University of Birmingham, UK (1998) in the field of Electrical and Electronics Engineering. His research interests are network security, cloud computing, Internet of Things (IoT) and non-linear wireless measurement system.



**Ahmad Fazli Abdul Aziz** graduated with Bachelor of Medicine and Bachelor of Surgery (MBBS) from the University of Sydney, Australia in 1999 and obtained MRCP (UK) in 2005. He is currently an Associate Professor of Medicine in University Putra Malaysia and a consultant cardiologist and electrophysiologist in Hospital Serdang, Malaysia.