Deep Multi-View Heartwave Authentication

C.L.P. Lim, W.L. Woo, S.S. Dlay, Di Wu and Bin Gao

***Abstract* — This paper presents a heartwave based authentication method that utilizes an ensemble of Deep Belief Networks (DBNs) under different parameters to increase the reliability of feature extraction. The multi-view outputs are further embed into a single view before inputting into a stacked DBN for classification. The result of the proposed novel architecture achieved a classification rate of 98.3% with 30% training data. Importantly, it is able to perform user classification using heartwave signals acquired under intense physical exercise where heartrate ranges from 50bpm to as high as 180bpm. Under extreme physical duress, the heartwave from individual experiences extreme morphological variations which render conventional classification approaches non-applicable.**

***Index Terms*—Deep Learning, Deep Belief Network, Discrete Wavelet Transformation, Multi-View Spectrum, Heartwave, Authentication.**

I. Introduction

THIS paper presents the use of individual heartwave signal as biometric mode to perform identification through a deep multiview ensemble learning methodology. With increasing evolution of digital technology, IoT solutions, cloud services and big data services, the need for secured data protection is univocal and many have implemented 2-Factor Authentication (2FA) similar to the security adopted by ebanking. Heartwave as a biometric mode has great potential to fulfil the security demands and ensuring access integrity. In addition, with increasing elderly population and longer life expectancy, elderly suffering from worn-out fingerprint and poor eyesight are facing challenges to use security system such as Digital Key Token and fingerprint biometric system comfortably. Equally, it is tormenting for elderly to setup password with periodic renewal and adhering to unique password characters combination. Heartwave as biometric mode has great potential to complement existing 2FA infrastructure for secured access to services and products through the means of wearable devices embedded with electrodes for heartwave signal acquisition. Apart from earlier example, heartwave signal as biometric mode can be used to enhance transportation safety such as authenticated access to vehicle with continuous monitoring of driver fatigue due to prolong driving[1, 2]. In healthcare, there are intense developments in tele-health solutions to provide continuously monitoring on the well-being of the elderly [3]. Biometric authentication for access to services enables medical personnel to respond to elderly needs reliably, securely and promptly. See Fig. 1 for illustration.

C.L.P. Lim, W.L. Woo, S.S. Dlay, Di Wu and Bin Gao are currently with School of Electrical and Electronic Engineering, Newcastle University, England, United Kingdom. Corresponding author e-mail: w.l.woo@ncl.ac.uk.

Unlike biometric modes of finger texture [4, 5], sclera [6, 7], iris [8-10], palm and face [11] or speech [12], heartwave signal does not require sophisticate setup for signal acquisition [13]. Unlike fingerprint where ridges can be worn out, heartwave signature is permanence. Heartwave signal can be acquired between two fingers of different hands which is the Lead I [14] signal from Bipolar Limb Leads group.

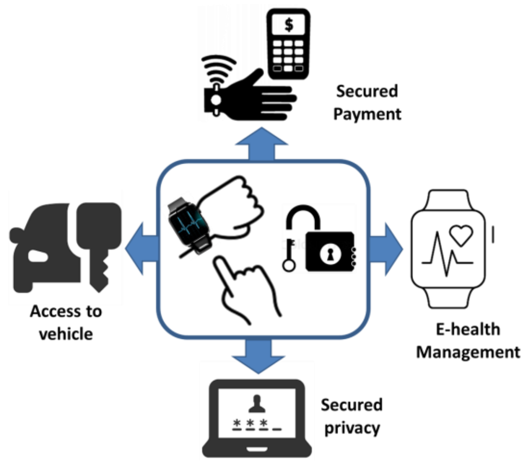


Fig. 1 Application scenarios of heartwave authentication

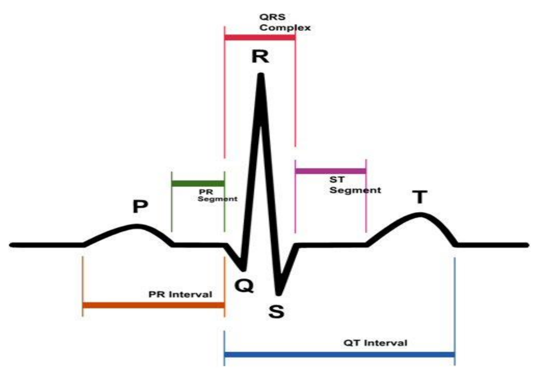


Fig. 2 Components of a heartwave signal

Heartwave signal comprises of three wave components namely P-Wave, QRS-Complex and T-Wave. In a single heartwave signal, it starts off with the contraction of the atrium muscle tissue which results in the formation of the P-Wave. Upon contraction, the excited electrical pulse travels to the ventricular muscle causing contraction which produces the QRS-Wave complex. QRS-Wave complex is the most

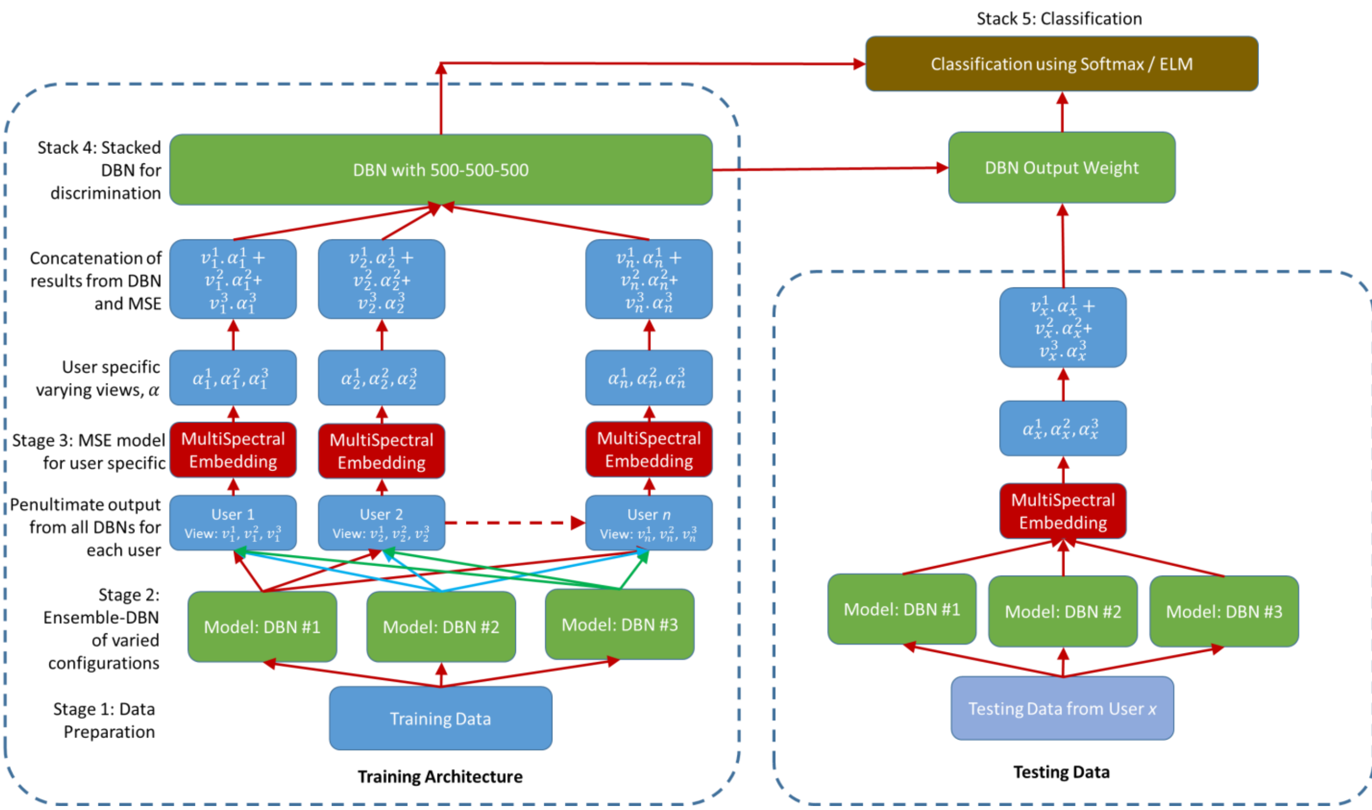


Fig. 3 Proposed architecture of ensemble-DBN coupled with MSE and stacked DBN

recognizable peak and is caused by the large muscular tissue mass at the ventricular segment. Following ventricular contraction, is relaxation of the ventricular which causes the formation of the T-Wave. See Fig. 2 for illustration of heartwave components. Individual heartrate is variable and the impetus of variation can be contributed by many factors such physiological activities, psychological and pathological related issues. While heartrate variation correlates greatly to the morphological variation of heartwave signal, the variation can also be caused by body embodiment. For example during resting, heartwave signal variation exists due to movement of the respiratory cage [15]. As heartrate of individual can vary as much as 400%, heartwave components P-Wave, QRS-Complex and T-Wave suffer from minimal to signification variation. This variation is one of the major dreaded challenges in inter and intra user authentication. The use of heartwave signal as biometric mode has aroused many research works with new and innovative approaches such as KNN classifiers (the most comment), LDA classifier[16], Support Vector Machine and Match Score Classifier [17, 18] and Generative Model Classifier [19-22]. Unfortunately, all the works use ECG data that are obtained under resting condition. Apart from statistical methodologies, there are works that use Neural Network approaches. The work in [23] uses neural network to identify and extract QRS complex which is use as a classifier in Deep Neural Net. While the result achieved an impressive 99.54% accuracy with a database of 90 individuals and used 70% of the data for training, it is to note that signals used in the studies were acquired under resting states. Along the same line, the works in [24, 25] use DNN to classify heartwave anomaly into 12 different categories of cardiac arrhythmia beats with an accuracy of 98.75%. To achieve the results, 90% of recordings are used for training and the remaining for testing. In actual implementation, it is not practical to use 90% of the data for training. Importantly, the data used in the training and testing are not acquired under physical duress.

Only one reported work by Agrafioti and Hatzinakos [26] uses data comprises of heartwave signal under varied conditions of health wellness. The work uses auto-correlation method to discard anomaly waveform of Premature Ventricular Contraction (PVC). PVC is a heart anomaly signal that occurs sporadically unlike the repetitive heartwave signal. Linear Discriminant is subsequently used to perform classification. Although under varied conditions, the work does not use signals that are acquired under physical duress. There are other published works for classifying heartwave anomaly due to cardiac related problem. In those works, heartwaves of multiple individuals are concatenated for anomaly detection. The works are not appropriate for use in biometric authentication.

This paper proposed a heartwave based authentication approach that is reliable and robust to varying heartrate. The pre-processing of heartrate signal incorporates heartrate dependent parameters to ensure reliable and accurate heartwave delineation. To ensure all features from varying heartrates are reliably extracted, ensemble of Deep Belief Networks (DBNs) configured with different parameters, are used to extract the features of heartwave. Thereafter the multi-view outputs from ensemble-DBN are combined into single view. Complementary properties from multi-view are embedded and subsequently input into a stack module of DBN to extract distinct and higher dimensional features for user classification. Extreme Learning Machine (ELM) is used at the final stage to provide fast and reliable authentication process.

II. DATABASE FOR PROPOSED WORK

In the testing and validation of the methodology, this work uses 2 public databases from Physionet. The first database comprises of 27 individuals. The heartwave signals from each individual are acquired under ECG treadmill. Each recording ranges from 20 minutes to as long as 40 minutes. Each of the users commenced from resting heartrate and progressed with increase intensity till the user either indicates their physical limits or when the maximum heartrate of the individual has been reached. Thereafter, the treadmill speed is throttled down to allow the individual to undergo a recovery period which is typically in the range of minimum 10 minutes. During this period, the heartwave signal is continuously recorded. Using this dataset, the heartwave of an individual undergoes maximum morphological changes in all components of the wave complexes namely, P-Wave, QRS-Complex and T-Wave.

As the dataset for heartwave signal under extreme physical duress is limited in quantity. Another dataset consisting of 25 users from Physionet is used. This dataset comprises of heartwave signal acquired under resting condition where the heartrate is fairly consistent without significant heartwave morphological variation. Each of the records from 25 users has a 5-minute duration. In total, more than 75,000 heartwaves are used in this development.

III. PROPOSED METHODOLOGY

The proposed method in this paper comprises of 5 stages. Stage 1 focuses on the data preparation prior for input to the ensemble-DBN. Stage 2 focuses on the use of ensemble-DBN as opposed to using single DBN for feature extraction. Stage 3 focuses on the combination of outputs from ensemble-DBN into a single view and the adoption of Multi-View Spectral Embedding method to achieve a single embedded output. Stage 4 focuses on the use of stacked DBN to support the eventual classification at Stage 5. The last stage focuses on the use an efficient method of Extreme Learning Machine (ELM) for accurate and reliable classification. Fig. 3 illustrates the proposed architecture.

A. Stage 1: Data Preparation

In the data preparation, the heartwave of each individual is independently delineated and extracted. The delineation process uses Discrete Wavelet Transform (DWT) with a fourth order Daubechies wavelet (DB4) to achieve heartwave delineation. However, DWT method has limitation for the delineation of heartwave from individual under elevated heartrate. Previously work has reported using DWT to extract heartwave of individual successfully [27]. However, it is to note that the mentioned work performed on signals acquired under non-physical duress condition.

Of the 52 heartwave records, 27 records are acquired under extreme physical duress where heartwave signal experienced maximum signal morphological changes. The heartrate of individual can change by as much as 400% where the wave components P-Wave, QRS Complex and T-Wave will experience compressed duration. T-Wave at elevated heartrate will experience morphological change in gradient slope as well.

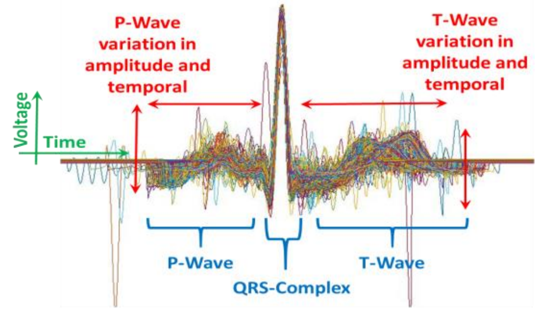


Fig. 4 Superimposition of all extracted heartwaves of an individual aligned around R-Peak. P-Wave and T-Wave show the morphological changes at varying heartrate.

To address the morphological changes, DWT is used to decompose the signal to the fifth level. The detection of each heartwave from individual commences from the detection of QRS-Complex followed by P-Wave and T-Wave. QRS-Complex comprises of the prominent R-Peak and detection is trivial using the typical peak detection function. To complete the delineation of heartwave individually, the detections of P-Wave and T-Wave require the reconstruction of DWT outputs using Approximate Coefficients of Level 1,2,3,4 and 5. The reconstruction exhibit prominent peaks and valleys to enable ease of detection. The selection of levels for reconstruction is highly dependent on the frequency components present in heartwave signal under varying heartrates.

Due to the presence of wide variations in heartrates, waves in particular to P-Wave and T-Wave experience significant morphological variations as shown in Fig. 4. At elevated heartrate, the duration of T-Wave can shrink by approximately 40%. To cater to the variation of P-Wave and T-Wave, two interval windows: PR and QT that are dependent on heartrate are defined. Once defined, the peaks and valleys of P-Wave and T-Wave on reconstructed signal are detected. The intervals of PR and QT are determined via (1) and (2) respectively where HR_{BPM} refers to the heartrate of an individual in beats-per-minute. The PR interval window is adopted from experimental studies conducted by [28-30]. The QT Duration is adopted using QT Nomogam [31, 32] which is a clinical risk assessment tool that predicts risk of QT prolongation in individual in respond to cardiac related drugs. See Fig. 5 and Fig. 6 for illustrations on PR and QT windows.

$$PR\ Interval\ (msec) = -0.351 \times HR_{BPM} + 176.7 \quad (1)$$

$$QT\ Duration\ (msec) = -2.2095 \times HR_{BPM} + 627.41 \quad (2)$$

Heartwave signals from all 52 individuals consisting of 75,188 heartwaves are successfully extracted. Of the 52 individuals, there are 5 individuals with inverted T-Waves. To further prepare the heartwaves for subsequent ensemble Deep Belief Network (DBN) training, each of the heartwave is aligned inside a fixed window of length 500 units. Using the location of the detected R-Peak, each heartwave is aligned at the 240th unit. The extremities of the extracted heartwave signal are padded with zeros to fill up the window length of 500 units. This provides a consistent data to the ensemble-DBN. See Fig. 7 for illustration.

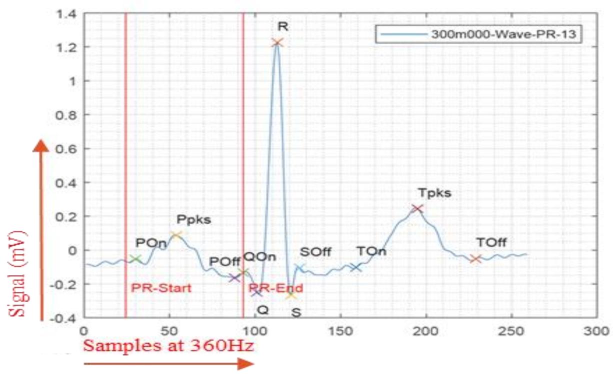


Fig. 5 PR Interval window for P-Wave extremity detection

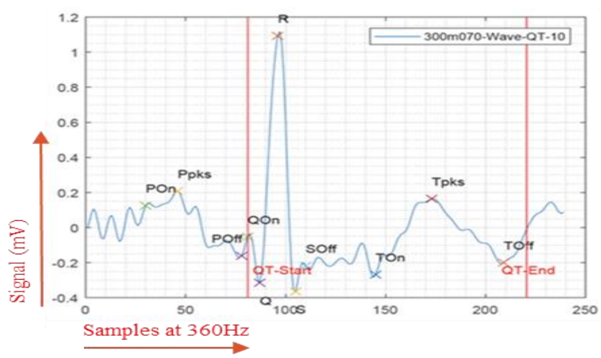


Fig. 6 QT Interval window for T-Wave extremity detection

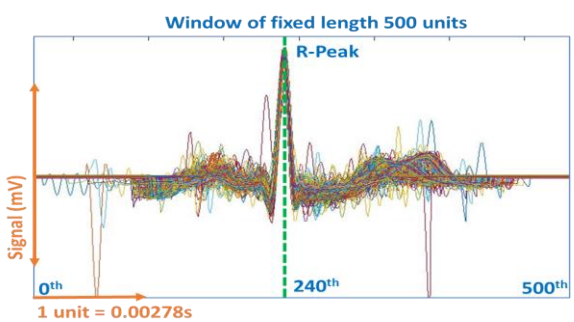


Fig. 7 Alignment of heartwave in a fixed window of length 500 units.

B. *Stage 2: Ensemble-Deep Belief Network (Ensemble-DBN)*

DBN is an undirected probabilistic model that is constructed by multiple layers of Restricted Boltzmann Machine (RBM). The approach to train a DBN is broadly split into two phases. The first phase deals with unsupervised pre-training followed by supervised back propagation fine-tuning. RBM is a shallow stochastic neural network comprises of one layer of visible units and one layer of hidden units. Characteristically, each visible unit is connected to all hidden units and vice versa. RBM is a bipartite graph and hence no visible unit is connected to any other visible unit and is equally said for hidden unit. See Fig. 8 for illustration. The theory of RBM based on the energy model is well established. Pre-training of unsupervised RBM is achieved by adopting a single step Contrastive Divergence (CD-1) to achieve the optimal parameters.

In a single DBN, three layers of RBM are stack onto each other. To achieve a supervised trained DBN, a classification layer consisting of 52 nodes is stacked above the third RBM. The output from the first pre-trained RBM becomes the visible input to the second RBM. Similarly pre-training continues till the third layer of RBM. Upon completion of pretraining for all the three layers of RBM, the pre-determined initial weights in the three layers are subjected to supervised training using backpropagation with fine tuning starting from the classification layer. Labelled training data are used in the supervised training.

To address signal noises and the morphological changes due to physical duress, 3 DBNs with different parameters are used to maximize features extraction.

The ensemble-DBN consists of the following configurations: DBN-1 with 500-100-500-52, DBN-2 with 500-500-500-52 and DBN-3 with 500-1000-500-52. Numeral 52 is the output layer for softmax based classification. The configuration of the ensemble-DBN including the number of DBNs are determined by hyper-parameters optimization using parameters of the number of hidden nodes, number of hidden layers and number of DBN based on the results of False Acceptance Rate and False Rejection Rate. The inputs data to the three DBNs are identical and the outputs from the penultimate layers of the ensemble-DBNs are subsequently input into Stage 3 for low dimensional embedding through multi-view spectral methodology.

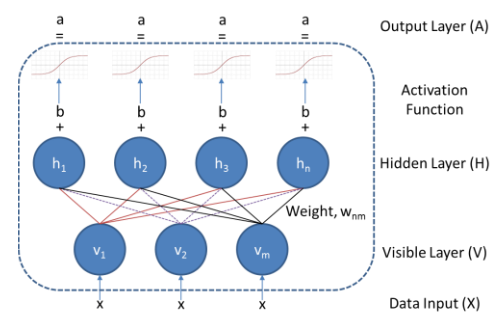


Fig. 8 A single hidden-layer RBM

C. *Stage 3: Multi-View Spectral Embedding*

Outputs of the penultimate layers of the ensemble-DBN contain the extracted features under different DBN configurations. Conversely, the outputs also contain features that are irrelevant and contribute to the misclassification rate in particular to low training data where the error in misclassification increases exponentially. The use of multi-view spectral embedding (MSE) is an effective approach to combine representations of ensemble-DBN through the identification of the complementary property of each view and embed to form a single view.

Prior to input to MSE, the randomized data is sort according to individual users. Hence, the data are sorted into 52-user specific datasets. Each of 52 datasets is individually input to MSE to determine the contribution factor α of each user specific view $\alpha = [\alpha_1, \dots, \alpha_m]$ where m refers to the number of DBNs in the ensemble-DBN.

Let X be the representation for each of the DBN input into the MSE module, $X = \{X^m \in \mathbb{R}^{m_i \times n}\}_{i=1}^{m}$ where n is the number of objects or samples. X^m is the feature matrix from each DBN. For each of the feature matrix,

$$X^i = [x_{1,}^i \dots . x_n^i] \in \mathbb{R}^{m_i \times n} \tag{3}$$

In each of the view, a patch is determined by considering an arbitrary point with its k-related samples using nearest neighbor. Hence, consider an arbitrary point x_j^i together with its k related points, the patch can be defined as $X_j^i = [x_{j,}^i x_{j1,}^i \dots . x_{jk}^i] \in \mathbb{R}^{m_i \times (k+1)}$. For each of the determined patch, there is a mapping and is given by $f_j^i : X_j^i \to Y_j^i$ where $Y_j^i = [y_{j,}^i y_{j1,}^i \dots . y_{jk}^i] \in \mathbb{R}^{d \times (k+1)}$ from each view, the part mapping is defined as (4):

$$\text{argmin}_{Y_j^i} \sum_{l=1}^{k} \left\| y_j^i - y_{jl}^i \right\|^2 \left(\omega_j^i\right)_l \tag{4}$$

where ω_j^i is a k-dimensional column vector weighted by $\left(\omega_j^i\right)_l = exp\left(-\left\| x_j^i - x_{jk}^i \right\|^2 / \gamma\right)$, γ defines the closeness of the neighborhoods. Therefore, the part optimization can be reformulated as (5):

$$\begin{aligned} &\underset{Y_j^i}{\text{argmin}}\, tr\left(\begin{bmatrix} \left(y_j^i - y_{jl}^i\right)^T \\ \dots \\ \left(y_j^i - y_{jk}^i\right)^T \end{bmatrix} \times \left[y_j^i - y_{jl}^i, \dots, y_j^i - y_{jk}^i\right] diag\left(\omega_j^i\right)\right) \\ &= \text{argmin}_{Y_j^i}\, tr\left(Y_j^i L_j^i \left(Y_j^i\right)^T\right) \end{aligned} \tag{5}$$

where $tr(\cdot)$ is the trace operator, $L_j^i = \begin{bmatrix} \sum_{l=1}^{k} \left(\omega_j^i\right)_l & -\left(\omega_j^i\right)^T \\ -\omega_j^i & diag\left(\omega_j^i\right) \end{bmatrix} \in \mathbb{R}^{(k+1)\times(k+1)}$ that encodes the complimentary property of a view into a low-dimensional embedding Y_j^i.

As MSE module is managing multiple views, there is a need to assign a weight factor to each of the represented views. This is to ensure that only complimentary property from multiple views contributes to the final embedment. To extract the complimentary property, nonnegative weights $\alpha = [\alpha_1, \dots, \alpha_m]$ are imposed on the part optimizations of each view. The significance of the complement property of a view is directly proportional to the value of α_i. Hence with the inclusion of weight to the represented m-th learned views, the summation of the jth part from the all views can be expressed as (6):

$$\underset{\left(Y=\left\{Y_j^i\right\}_{i=1}^{m}, \alpha\right)}{\text{argmin}} \sum_{i=1}^{m} \alpha_i tr\left(Y_j^i L_j^i \left(Y_j^i\right)^T\right) \tag{6}$$

To ensure all patches are reference to global origin, all Y_j^i can be mapped together with the assumption that the coordinate for $Y_j^i = [y_{j,}^i y_{j1,}^i \dots . y_{jk}^i]$ is selected from the global coordinate $Y = [y_1, y_2, y_3, \dots, y_n]$, i.e., $Y_j^i = Y S_j^i$, where $S_j^i \in \mathbb{R}^{n \times (k+1)}$ is the selection matrix to encode the spatial relationship of patch samples from the original high-dimensional space. Therefore, (5) can be rewritten as (7):

$$\underset{Y_j, \alpha}{\text{argmin}} \sum_{i=1}^{m} \alpha_i^{\varepsilon} tr\left(Y S_j^i L_j^i \left(S_j^i\right)^T (Y)^T\right) \tag{7}$$

By summing all part optimizations, the global coordinate alignment is given by (8):

$$\begin{cases} \underset{Y, \alpha}{\text{argmin}} \sum_{i=1}^{m} \alpha_i^{\varepsilon} tr(Y L^i Y^T) \\ s.t. YY^T = I, \sum_{i=1}^{m} \alpha_i = 1, \alpha_i \geq 0 \end{cases} \tag{8}$$

where $L^i \in \mathbb{R}^{n \times n}$ is the alignment matrix for the mth learned representations, and it is defined as $L^i = \sum_{j=1}^{n} S_j^i L_j^i \left(S_j^i\right)^T$. The constraint $YY = I$ is to uniquely determine Y. Exponent ε is the coefficient for controlling the interdependency between different views where $\varepsilon \geq 1$. By performing a normalization on L^i, we obtain a normalized graph Laplacian L_{sys} which is symmetric and positive semidefinite as defined in (9):

$$L_{sys} = D^{i^{-\frac{1}{2}}} L^i D^{i^{-\frac{1}{2}}} = I - D^{i^{-\frac{1}{2}}} Q^i D^{i^{-\frac{1}{2}}} \tag{9}$$

where $Q^i \in \mathbb{R}^{n \times n}$ and $[Q^i]_{lj} = exp\left(-\left\| x_n^i - x_j^i \right\|^2 / \gamma\right)$ if x_n^i is among the k-nearest neighbors of x_j^i or vice versa; $[Q^i]_{lj} = 0$ otherwise. D^i is a degree matrix defined as the diagonal matrix with the degrees $[D^i]_{jj} = \sum_l [Q^i]_{lj}$. Equation (8) is a nonlinearly constrained nonconvex optimization problem and the optimal solution can be obtained by using Expectation Maximization (EM). Iteratively, the optimization of Y and α can be determined.

M-step: Fix Y to update α.

By introducing a Lagrange multiplier λ and to take the constraint $\sum_{i=1}^{m} \alpha_i = 1$ into consideration, the Lagrange function can be expressed as (10):

$$L(\alpha, \lambda) = \sum_{i=1}^{m} \alpha_i^{\varepsilon} tr\left(Y L_{sys} Y^T\right) - \lambda\left(\sum_{i=1}^{m} \alpha_i - 1\right) \tag{10}$$

The solution for α_i can be given by (11):

$$\alpha_i = \frac{\left(1/tr\left(YL_{sys}Y^T\right)\right)^{1/(\varepsilon-1)}}{\sum_{i=1}^{m}\left(1/tr\left(YL_{sys}Y^T\right)\right)^{1/(\varepsilon-1)}} \qquad (11)$$

When Y is fixed, (11) gives the global optimal α.

E-step: Fix α to update R.

The optimization problem in (8) is equivalent to (12):

$$\min_R(YLY^T) \quad s.t. YY^T = I \qquad (12)$$

where $L = \sum_{i=1}^{m} \alpha_i^{\varepsilon} L_{sys}$. Based on Ky-Fan theorem, (8) has a global optimal solution when α is fixed. The optimal R is given as the eigenvectors associated with the smallest d eigenvalues of the matrix L.

With the optimized α_i obtained, it is multiplied to the penultimate layers of the respective views and summed. This ensures that for each user, only significant complementary property is amplified. This process is applies to all users.

D. Stage 4: Ensemble-DBN with Stacked DBN

At this stage, a second DBN is stacked above the MSE module for extraction of higher order features. The input to the stacked DBN can be represented as (13):

$$\Psi_m = f_{MSE}\left(D(\sigma_i) + \cdots + \mathrm{D}(\sigma_m)\right) \qquad (13)$$

where f_{MSE} is the output of MSE, i refers to the number of DBN in the ensemble-DBN and σ_i refers to the product of α_i and the respective penultimate layer of the ensemble-DBN. With the contribution of weightage from MSE, the view with significant domain features will provide greater learning probability to the DBN. This is in contrast to linear combination where the inputs from the three views are averaged which leads to significant domain features being suppressed and lesser significant domain features being elevated. The DBN used in this stack is configured with the following configuration of 500-500-500-52. The pretraining and supervised training for the stack DBN is similar to the ensemble-DBN.

E. Stage 5: Classification using Extreme Learning Machine

From the penultimate layer of the stacked DBN, Extreme Learning Machine (ELM) method is used as a classifier. The advantage of ELM over Softmax is the significant lower computation speed required to train and test the samples. Another difference between ELM and Softmax lies on the training process. Softmax is an iterative process to fine tune the weights in the classification layer. ELM however is an analytical process that relies on universal approximation. ELM is essentially a single hidden-layer feedforward neural network (SLFN). ELM accomplishes through the assignment of random weights and biases to the hidden nodes and subsequently uses matrix computations to determine the output weights. Given N as inputs to ELM where $\{s_{i,}t_i\}_{i=1}^{N}$. ELM model with K hidden nodes can be written as defined in (14):

$$t_i = \sum_{\kappa=1}^{K} g(s_i, u_\kappa, v_\kappa)\,\beta_\kappa, \quad i = 1, \ldots, N \qquad (14)$$

where t_i is the output, x_i denote the input vector, u_κ and v_κ are the parameters of the activation function of the κth hidden node, $g(x_\phi, u_\kappa, v_\kappa)$ is the output of the κth hidden node with respect to the κth input. β_κ is the output weight of the κth hidden node.
The expression of (14) can be re-written as (15)

$$\boldsymbol{T} = \boldsymbol{S\beta} \qquad (15)$$

where $\boldsymbol{T} = [t_1, \ldots, t_i]^T$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]^{\boldsymbol{T}}$, and the hidden output matrix (16)

$$\boldsymbol{S} = \begin{bmatrix} s(x_1, u_1, v_1) & \cdots & s(x_1, u_K, v_K) \\ \vdots & \ddots & \vdots \\ s(x_N, u_1, v_1) & \cdots & s(x_N, u_K, v_K) \end{bmatrix}_{N\times K} \qquad (16)$$

An ELM learns the parameters in two stages: 1) random feature mapping and 2) linear parameter solving. In the first stage, the input data is projected into a feature space with randomly initialized parameters using the activation function. It has been shown that the randomly initialized parameters are able to approximate any continual function [33, 34]. Therefore, the only parameter that needs to be determined is the output weight $\boldsymbol{\beta}$, which can be estimated by (17)

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{S}^{\dagger}\boldsymbol{T} \qquad (17)$$

where $\boldsymbol{S}^{\dagger}$ is the Moore-Penrose generalized inverse. The use of ELM offers the benefits that have been well documented by [35, 36]. Advantages include local minimal, overtraining and significantly lower computing resources. Due to single matrix operation, it leads to extremely efficient computation.

IV. Experimental results

In the validation of classification performance on the proposed architecture, heartwave from 52 individuals are used. Of the 52 individuals, 5 individuals have heartwaves which are non-healthy. To determine the accuracy of the performance, performance indicator Positive Predictive Value (PPV) of the following form is used:

$$PPV = \frac{No. of\ True\ Positive}{No.\ of\ True\ Positive + No.\ of\ False\ Positive} \qquad (18)$$

where True Positive refers to the ability to classify the individual correctly and False Positive refers to an invalid classification.

A. Classification with Anomaly Heartwave Signal

Heartwave based biometric methodologies have been actively tested on individual with normal heartwave signals. Conversely, there are also work in classifying unhealthy heartwave signals into different categories [24, 37]. However, there has been no investigation on the possibility of including individual with anomaly heartwave as an authentication subject. For some individuals, the anomaly heartwave signature occurs in every heartwave such as the extended T-Peak. For others, the anomaly can be due to the abnormal R-R interval where the period between R peaks are not consistent. See Fig. 9 and Fig. 10 for illustrations.

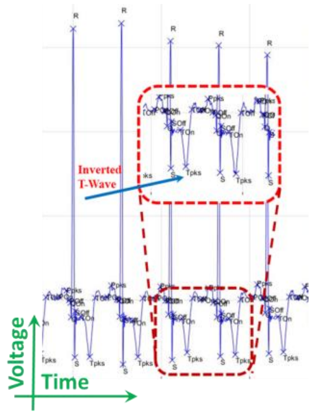


Fig. 9 Anomaly signal with inverted T-Wave that occurs in every heartwave and anomaly signal with extended T-Peak and inconsistent R-R interval

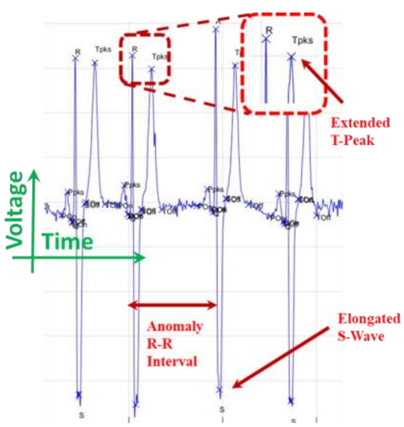


Fig. 10 Anomaly signal with extended T-Peak and inconsistent R-R interval

The proposed architecture is tested with a mixture of healthy and unhealthy heartwave signals. This test determines the possibility of individual with anomaly signal having unique characteristic to be authenticated.

The proposed architecture is tested on two separate datasets in which one of the dataset contains normal heartwave and the second dataset contains anomaly heartwave signal. Lastly, a third dataset containing both normal and anomaly heartwave is generated. The performance results are tabulated in Table I.

From the test performance as shown in Table I, misclassification rate of 0.8% is achieved on dataset contained healthy heartwave signal. A misclassification rate of 1.8% is achieved on dataset that contains anomaly heartwave signal. The dropped in performance is likely due to the limited data available for training. In combining all heartwave signal into a single dataset, the overall performance achieved a misclassification rate of 1.1%. In contrast to a statistical method of using Gaussian Mixture Model with Hidden Markov Model (GMM-HMM), the misclassification rate achieved is 25% under similar proportional of training and testing data. In the statistical based architecture, it uses the characteristic strength of GMM and HMM for feature clustering and classification. For each individual, the dataset belonging to a single individual is clustered using GMM. This clustering of individual dataset using GMM allows a user representative model to be developed. The GMM is subsequently used by HMM to develop into a user specific HMM model to support classification.

Table I. Proposed architecture performance on normal and anomaly heartwave signal

		Misclassification Results (%)
Misclassification Rates on Categories of Heartwave Signal	Proposed Architecture on Normal Heartwave	0.71
	Proposed Architecture on Anomaly Heartwave	1.69
	Proposed Architecture Combine (Normal + Anomaly)	1.07
	GMM-HMM (Combine: Normal + Anomaly)	**25.00**

The test has shown the possibility of inclusion of individual with anomaly heartwave in using heartwave as a biometric mode.

B. Performance of Single DBN vs Ensemble-DBN

The proposed architecture uses 3 signal DBNs and combined the outputs from each DBN before transferring to single stack DBN.

There are reported works that use Deep Learning in biometric authentication and only single DBN is adopted. Primarily, the dataset used in their works are homogenous and most importantly, heartwave signals are acquired under non-physical duress.

In the process of optimizing DBN configuration, the hyper parameters consisting of hidden nodes and layers are used to determine the optimized DBN configurations. Considering the dataset that consists of heartwave signals acquired under physical duress, the Signal-to-Noise (SNR) of heartwave signals varies greatly within and between individuals. This is primarily due to the placement of the electrodes, motion artifact and heartrate. Fig. 11 illustrates a low SNR of a noisy heartwave signal. Further investigations reveal that for heartwave signal with high SNR, the DBN configuration with lower number of hidden nodes performs well in classification, in excess of 98% accuracy. Conversely, for heartwave signal with low SNR, DBN with higher number of hidden nodes performs better in classification. To further complicate the classification process, the SNR of most individual heartwave signals varies at different heartrates. During the heartwave signal acquisition process, for most individual, at the early stage of the treadmill testing where the heartrate is near resting heartrate, the signal has high SNR. With increasing intensity of the treadmill, the SNR deteriorates at increasing heartrates. This concludes that having multiple DBNs is necessary to manage the varying SNR of individual heartwave signal.

In this investigation, the performance of the proposed configuration for each of the three DBNs are compared. The results are shown in Table II. For single DBN with configuration of 500-100-500, the overall classification accuracy is in excess of 91% as compared to DBNs with higher number of hidden nodes. This is primarily because of the large number of heartwaves having higher SNR at lower heartrate. Leveraging on the strength of each of the single DBN, the three outputs are combined through the adoption of MSE. The MSE determines view with higher significant strength which output respective weight for each of the views. The greater the significance of the view, the higher the weight. This results in the ability of the proposed architecture to achieve a classification accuracy of 98.3%.

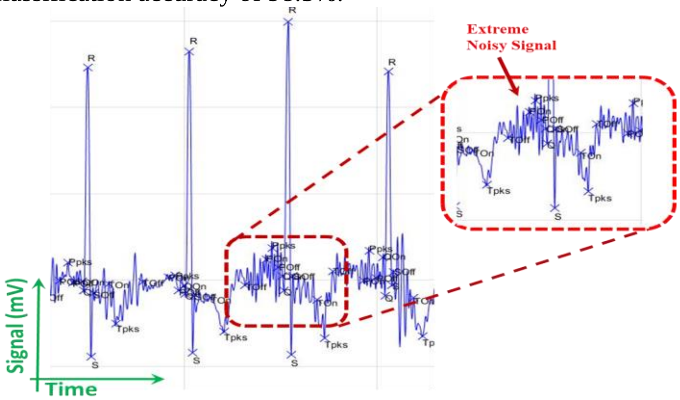


Fig. 11 Extreme noisy signal with low SNR

Table II. Classification performance of various standalone DBN against proposed architecture

		Classification Result (%)
	Proposed Architecture	**0.99**
Stage 2 DBNs	Stage2: DBN#1 @ 500x500x500	0.9
	Stage2: DBN#2 @ 500x1000x500	0.88
	Stage2: DBN#3 @ 500x100x500	0.92

C. Classification Performance on Proportion of Training

In some of the reported works, the proportion of training data against testing data is benchmarked from 70% to 90% [23, 24]. While having the 70% of the data may seem appropriate, it is necessary to explore the strength of Deep Learning with the aim to reduce the training data required. Importantly, it is of the opined that the apportioning of 70% of data for training is still excessive.

The investigation starts with the varying percentage of training data. In addition, the investigation also examines the performance of the three single DBNs against the proposed architecture at different proportion of training data. The results of the classification are summarized in the Fig. 12. From Fig. 12, it is evident that the conventional apportioning of 70% of data for training is reasonable. At 70% of data for training, the classification performance easily achieved a classification accuracy in excess of 95%. With decreasing training data, the performance deteriorates rapidly.

For the proposed architecture, the classification performance remains consistent at about 98.3% even at 30% training data. This is in stark contrast to the performance of individual DBNs. As discussed in Section IV-B, where due to variable heartrate, it leads to varying signal-to-noise ratio. Each of the DBN has limited capability to capture all the morphological variations of heartwave signal. For heartwave signal with high SNR, the DBN configuration with lower number of hidden nodes performs differently in contrast to heartwave signal with low SNR. Importantly, this reinforces the importance of having multiple DBNs with varying configurations and incorporating MSE into the architecture.

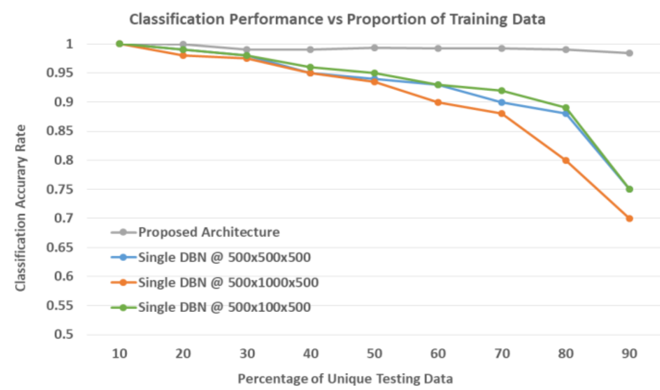


Fig. 12 Comparison of classification performance at different portion of training data

D. Comparison of Proposed Method with Statistical Methodology

The performance of the proposed architecture is evaluated and compared against other statistical methodologies. The work in [38] uses a combination of fiducial, non-fiducial data of heartwave signal and inertial sensor parameter and achieved a True Positive Rate of approximately 80% with False Negative Rate of 12%. Another commonly adopted methodology is the generative modelling such as Gaussian Mixture Modelling and Hidden Markov Modelling which are used to support data clustering and classification. Hence, the proposed architecture is compared against generative modelling of Gaussian Mixture Model with Hidden Markov Model. The GMM-HMM method achieved a True Positive Rate of approximately 90% with False Negative Rate of 10%. In comparison of proposed architecture, the True Positive Rate achieved is 98% with False Negative Rate of under 2%. See Table III for comparison. Deep learning methodology offers a highly reliable approach to classification.

Table III. Comparison of classification performance between proposed architecture and statistical methods

		Results	
		True Positive Rate, %	**False Positive Rate, %**
ROC Performance of Different Methodologies	**Proposed Methodology**	**98**	**2**
	GMM-HMM	90	10
	Inertial Sensor Approach	80	12

E. Performance of Proposed Method with Architecture 2

The proposed architecture uses a coefficient factor from MSE to amplify view with significant contribution and the expression (12) can be re-expressed into the following form where v_i represents the penultimate layer of DBN i and α_i refers to the multiplier factor for view i from MSE.

$$\Psi_{user} = f_{MSE}(v_i\alpha_i + \cdots + v_m\alpha_m) \quad (19)$$

An alternative architecture (termed as Architecture 2) has been investigated. In Architecture 2, all the outputs of the penultimate layers are arithmetically sum and average. The output of the MSE, Y, which contains the embedded low-dimensional complementary property from all the views is concatenate with the averaged penultimate layers. The input to the stacked DBN can be expressed as shown in (28). The eventual dataset, Ψ_{user} from Architecture 2 is of higher dimension than proposed architecture due to the concatenation.

$$\Psi_{user} = f_{MSE}\left(Concat\left(\frac{v_i + \cdots + v_m}{m}, Y\right)\right) \qquad (20)$$

Fig. 13 shows the performance similarity between the proposed architecture and the alternative architecture. The performance between the proposed and alternative are similar at approximately around 98.2% even at reduce training sample size.

The performance of Architecture 2 reaffirms the strength of MSE. The difference between expression (19) and (20) is the used of MSE output. The proposed architecture uses multiplier factor α, to amplify view with significant contribution. Architecture 2 uses Y, which is the encoded low dimensional data structure contributed by MSE. Although the multiplier factor is not utilized, the encoded low dimension structure Y has already been encoded with views containing significant property. Thus the encoded low dimensional structure plays a key role in the higher module classification.

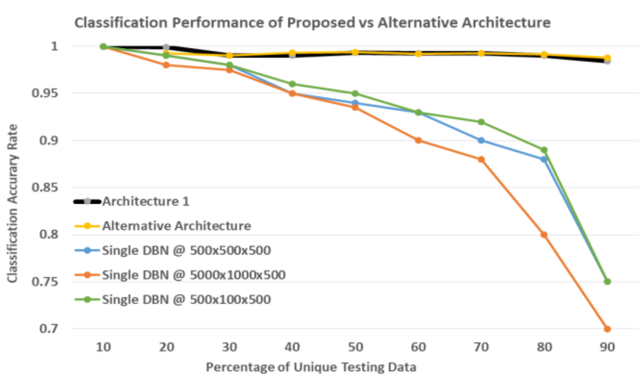


Fig. 13 Comparison of classification performance between proposed architecture and architecture 2 (alternative)

V. Conclusion

The proposed architecture in the heartwave based biometric authentication exceeded statistical methodology. Importantly, classification on individual heartwave with intense varying heartrate which causes signification morphologically variations has been tested by the proposed architecture. The proposed architecture has shown the ability to identify individuals comprising of normal and abnormal heartwave signals with high level of reliability. Architecturally, the ensemble-DBN is necessary to enable feature extractions under different morphological variations. The incorporation of MSE has enabled views containing significant features with greater biasness in aggregated inputs to the stacked DBN module. While the proposed architecture has proven to be successful in classification, better performance can be expected with more data in particular to the heartwave signal acquired under intense physical duress and abnormally heartwave.

Computing authentication features using the proposed deep multi-view authentication is computationally intensive. In the near term development, apart from acquiring more data to validate the methodology, there is a need to investigate the comprehensiveness and suitability of using heartwave signal acquired from the both fingers to support authentication. Concurrently, it is also important to establish a framework in engaging cloud based services to support training and trial of authentication.

VI. References

[1] K. T. Chui, K. F. Tsang, H. R. Chi, B. W. K. Ling, and C. K. Wu, "An Accurate ECG-Based Transportation Safety Drowsiness Detection Scheme," *IEEE Transactions on Industrial Informatics,* vol. 12, pp. 1438-1452, 2016.

[2] C. K. Wu, K. F. Tsang, and H. R. Chi, "A wearable drunk detection scheme for healthcare applications," in *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, 2016, pp. 878-881.

[3] G. Yang, L. Xie, M. Mäntysalo, X. Zhou, Z. Pang, L. D. Xu*, et al.*, "A Health-IoT Platform Based on the Integration of Intelligent Packaging, Unobtrusive Bio-Sensor, and Intelligent Medicine Box," *IEEE Transactions on Industrial Informatics,* vol. 10, pp. 2180-2191, 2014.

[4] Raid Al-Nima, Mohammed Abdullah, S. S. Dlay, W. L. Woo and J. A. Chambers, "Finger Texture Biometric Verification Exploiting a Multi-scale Sobel Angles Local Binary Pattern and Score-based Fusion," *Digital Signal Processing*, vol. 70, pp. 178-189, 2017.

[5] R.R.O. Al-Nima, S. S. Dlay, Al-Sumaidaee, W. L. Woo and J. A. Chambers,, "Robust Feature Extraction and Salvage Schemes for Finger Texture Based Biometrics," *IET Biometrics*, vol. 6, no. 2, pp. 43-52, 2016.

[6] Sinan Alkassar, W.L. Woo, S.S. Dlay, and Jonathon Chambers, "Sclera Recognition: On The Quality Measure and Segmentation of Degraded Images Captured Under Relaxed Imaging Conditions," IET Biometrics, vol. 6, no. 4, pp. 266–275, 2017.

[7] Sinan Alkassar, W.L. Woo, S.S. Dlay, and Jonathon Chambers, "Robust Sclera Recognition System with Novel Sclera Segmentation and Validation Techniques," *IEEE Trans. on Systems, Man and Cybernetics: Systems*, no. 47, no. 3, pp.474-486, 2017.

[8] Mohammed Abdullah, S.S. Dlay, W.L. Woo and J.A. Chamber, "A Novel Framework for Cross-Spectral Iris Matching," *IPSJ Transactions on Computer Vision and Applications*, vol. 8, no. 9, 2016.

[9] Mohammed Abdullah, S.S. Dlay, W.L. Woo, and Jonathan Chambers, " Robust Iris Segmentation Method Based on a New Active Contour Force with a Non-ideal Normalization," *IEEE Trans. on Systems, Man and Cybernetics: Systems*, no. 47, no. 12, pp.3128-3141, 2016.

[10] M.A.M Abdullah, S.S. Dlay, W.L. Woo, J.A. Chambers, "A framework for iris biometrics protection: A marriage between watermarking and visual cryptography," *IEEE Access*, vol. 4, pp. 10180-10193, 2016.

[11] M.I. Muhammad, S.S. Dlay and W.L. Woo, " Non-Stationary Features Fusion for Face and Palmprint Multimodal Biometrics," *Neurocomputing*, vol. 177, pp. 49-61, 2016.

[12] Musab Al-Kaltakchi, W.L. Woo S.S. Dlay and J.A. Chambers, "Evaluation of a Speaker Identification System With and Without Fusion Using Three Databases in the Presence of Noise and Handset Effects," *EURASIP Journal on Advances in Signal Processing*, 2017:80 (https://doi.org/10.1186/s13634-017-0515-7)

[13] Y. Si, J. Mei, and H. Gao, "Novel Approaches to Improve Robustness, Accuracy and Rapidity of Iris Recognition Systems," *IEEE Transactions on Industrial Informatics,* vol. 8, pp. 110-117, 2012.

[14] A. Lourenço, H. Silva, and A. Fred, "Unveiling the Biometric Potential of Finger-Based ECG Signals," *Computational Intelligence and Neuroscience,* vol. 2011, p. 8, 2011.

[15] G. R. Shaw and P. Savard, "On the detection of QRS variations in the ECG," *IEEE Trans Biomed Eng,* vol. 42, pp. 736-41, Jul 1995.

[16] K. A. Sidek, I. Khalil, and M. Smolen, "ECG biometric recognition in different physiological conditions using robust normalized QRS complexes," in *Computing in Cardiology (CinC), 2012*, 2012, pp. 97-100.

[17] K. Kyeong-Seop, Y. Tae-Ho, L. Jeong-Whan, K. Dong-Jun, and K. Heung-Seo, "A Robust Human Identification by Normalized Time-Domain Features of Electrocardiogram," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2005, pp. 1114-1117.

[18] A. V. Lyamin and E. N. Cherepovskaya, "An evaluation of biometrie identification approach on low-frequency ECG signal," in *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2017, pp. 000137-000142.

[19] P. Shing-Tai, W. Yi-Heng, K. Yi-Lan, and C. Hung-Chin, "Heartbeat Recognition from ECG Signals Using Hidden Markov Model with Adaptive Features," in *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2013 14th ACIS International Conference on*, 2013, pp. 586-591.

[20] E. Rabhi and Z. Lachiri, "Biometric personal identification system using the ECG signal," in *Computing in Cardiology Conference (CinC), 2013*, 2013, pp. 507-510.

[21] P. Shing-Tai, H. Tzung-Pei, and C. Hung-Chin, "ECG signal analysis by using Hidden Markov model," in *Fuzzy Theory and it's Applications (iFUZZY), 2012 International Conference on*, 2012, pp. 288-293.

[22] W. Louis, S. Abdulnour, S. J. Haghighi, and D. Hatzinakos, "On biometric systems: electrocardiogram Gaussianity and data synthesis," *EURASIP Journal on Bioinformatics and Systems Biology,* vol. 2017, p. 5, February 21 2017.

[23] A. Page, A. Kulkarni, and T. Mohsenin, "Utilizing deep neural nets for an embedded ECG-based biometric authentication system," in *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2015, pp. 1-4.

[24] Y. Yan, X. Qin, Y. Wu, N. Zhang, J. Fan, and L. Wang, "A restricted Boltzmann machine based two-lead electrocardiography classification," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 2015, pp. 1-9.

[25] P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T. s. Zhou*, et al.*, "High-Performance Personalized Heartbeat Classification Model for Long-Term ECG Signal," *IEEE Transactions on Biomedical Engineering,* vol. 64, pp. 78-86, 2017.

[26] F. Agrafioti and D. Hatzinakos, "ECG biometric analysis in cardiac irregularity conditions," *Signal, Image and Video Processing,* vol. 3, pp. 329-343, 2008.

[27] C. L. P. Lim, W. L. Woo, and S. S. Dlay, *Enhanced Wavelet Transformation for Feature Extraction in Highly Variated ECG Signal*: Institution of Engineering and Technology, 2015.

[28] S. G. Carruthers, B. McCall, B. A. Cordell, and R. Wu, "Relationships between heart rate and PR interval during physiological and pharmacological interventions," *British Journal of Clinical Pharmacology,* vol. 23, pp. 259-265, 1987.

[29] "PR Interval Behavior During Exercise Stress Test FAU - Lee, Jae Ung FAU - Kim, Kyung Soo FAU - Kim, Jeong Hyun FAU - Lim, Heon Kil FAU - Lee, Bang Hun FAU - Lee, Chung Kyun," *Korean J Intern Med,* vol. 10, pp. 137-142, 7 1995.

[30] J.-H. Atterhög and E. Loogna, "P-R interval in relation to heart rate during exercise and the influence of posture and autonomic tone," *Journal of Electrocardiology,* vol. 10, pp. 331-336, // 1977.

[31] A. Chan, G. K. Isbister, C. M. Kirkpatrick, and S. B. Dufful, "Drug-induced QT prolongation and torsades de pointes: evaluation of a QT nomogram," *Qjm,* vol. 100, pp. 609-15, Oct 2007.

[32] A. A. Fossa, T. Wisialowski, A. Magnano, E. Wolfgang, R. Winslow, W. Gorczyca*, et al.*, "Dynamic Beat-to-Beat Modeling of the QT-RR Interval Relationship: Analysis of QT Prolongation during Alterations of Autonomic State versus Human Ether a-go-go-Related Gene Inhibition," *Journal of Pharmacology and Experimental Therapeutics,* vol. 312, pp. 1-11, January 1, 2005 2005.

[33] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neural Computing,* vol. 70, pp. 489-501, Dec 2006.

[34] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 42, pp. 513-529, Apr 2012.

[35] A. R. Hassan, "A comparative study of various classifiers for automated sleep apnea screening based on single-lead electrocardiogram," in *2015 International Conference on Electrical & Electronic Engineering (ICEEE)*, 2015, pp. 45-48.

[36] L. L. C. Kasun, Y. Yang, G. B. Huang, and Z. Zhang, "Dimension Reduction With Extreme Learning Machine," *IEEE Transactions on Image Processing,* vol. 25, pp. 3906-3918, 2016.

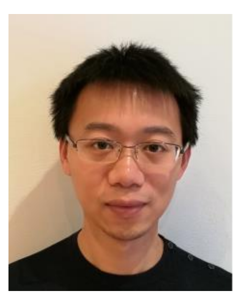
[37] M. Huanhuan and Z. Yue, "Classification of Electrocardiogram Signals with Deep Belief Networks," in *2014 IEEE 17th International Conference on Computational Science and Engineering*, 2014, pp. 7-12.

[38] J. C. Sriram, M. Shin, T. Choudhury, and D. Kotz, "Activity-aware ECG-based patient authentication for remote health monitoring," presented at the Proceedings of the 2009 international conference on Multimodal interfaces, Cambridge, Massachusetts, USA, 2009.

C.L.P. Lim received his B.Eng (Hons) degree in Mechatronics Engineering from University of New South Wales, Australia in 2000. In 2009, he obtained his Dual M.Sc. degrees (Micro Electro Mechanical System, MEMS) from Nanyang Technological University, Singapore and École Supérieure d'Ingénieurs en Électrotechnique et Électronique (ESIEE), France. He is currently working towards his PhD degree at Newcastle University focusing on heartwave based biometric authentication using methodologies of statistical tools and neural network.

W.L. Woo (M'09–SM'11) received the BEng degree in Electrical and Electronics Engineering and the PhD degree from the Newcastle University, UK. He was awarded the IEE Prize and the British Scholarship. He is currently a Reader in Intelligent Signal Processing and Director of Operations for the international branch of the university in Singapore. His major research is in the mathematical theory and algorithms for machine learning and multidimensional signal processing. Dr. Woo is a Member of the Institution Engineering Technology. He was a recipient of the IEE Prize and the British Scholarship. Currently, he serves on the editorial board of the several international signal processing journals.

S.S. Dlay received his BSc (Hons.) degree in Electrical and Electronic Engineering and his PhD in VLSI Design from the Newcastle University. During this time he held a Scholarship from the Engineering and Physical Science Research Council (EPSRC) and the Charles Hertzmann Award. In 1984 he was appointed as a Lecturer in the Department of Electronic Systems Engineering at the University of Essex. In 1986 he re-joined the Newcastle University as a Lecturer, then in 2006 he was appointed to a Personal Chair in Signal Processing Analysis. Professor Dlay is a College Member of the EPSRC.



Di Wu received his B.S. degree from Minzu University of China, Beijing, China in 2010, the M.Sc. degree in communications and signal processing and PhD degree from Newcastle University, Newcastle upon Tyne, U.K. (2011-2017). His research interests include deep learning, audio and image processing, source separation, and object recognition.

Bin Gao (M' 12-SM' 14) received his B.S. degree in communications and signal processing from Southwest Jiao Tong University (2001-2005), China, MSc degree in communications and signal processing with Distinction and PhD degree from Newcastle University, UK (2006-2011). His research interests include sensor signal processing, machine learning, nondestructive testing and evaluation. Currently, he is a Professor with the School of Automation Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu, China.