IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 16, 2018, accepted May 15, 2018, date of publication May 24, 2018, date of current version June 20, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2838559

Adaptive Hybrid Feature Selection-Based Classifier Ensemble for Epileptic Seizure Classification

FARRIKH ALZAMI[1], JUAN TANG[2], ZHIWEN YU[1], (Senior Member, IEEE), SI WU[1], C. L. PHILIP CHEN[3], (Fellow, IEEE), JANE YOU[4], AND JUN ZHANG[1], (Fellow, IEEE)

[1]School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
[2]Guangzhou Women and Children's Medical Centre, Guangzhou 510623, China
[3]Department of Computer and Information Science, University of Macau, Macau 99999, China
[4]Department of Computing, The Hong Kong Polytechnic University, Hong Kong

Corresponding authors: Farrikh Alzami (alzami@dsn.dinus.ac.id) and Juan Tang (moyudaoyuan@163.com)

This work was supported in part by NSFC under Grant 61722205, Grant 61751205, Grant 61572199, Grant 61572540, Grant 61472145, and Grant U1611461, in part by the Guangdong Natural Science Funds under Grant S2013050014677 and Grant 2017A030312008, in part by the Science and Technology Planning Project of Guangdong Province, China, under Grant 2015A050502011, Grant 2016B090918042, Grant 2016A050503015, and Grant 2016B010127003, in part by the Guangzhou Science and Technology Planning Project under Grant 201704030051, in part by the Macau Science and Technology Development under Grant 019/2015/A and Grant 024/2015/AMJ, in part by the Multiyear Research Grants from University of Macau Multiyear Research Grants, in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Grant CityU 11300715, in part by the Hong Kong General Research Grant GRF, under Grant 152202/14E, in part by the PolyU Central Research Grant (G-YBJW), and in part by the City University of Hong Kong under Project 7004884.

ABSTRACT Feature selection and ensemble learning can be used to improve the accuracy and robustness of epileptic seizure detection and classification. Unfortunately, a few studies have fully utilized feature selection and ensemble learning. In this paper, we present an adaptive hybrid feature selection-based classifier ensemble (AHFSE) for epileptic seizure classification. The AHFSE creates new sample subsets in every bootstrap using adaptive hybrid feature selection. It combines them using rank aggregation to obtain a distinguished subset of features. These new samples' subsets are then fed into a classifier. Finally, majority voting is used to complete the detection and classification tasks. The AHFSE is designed to obtain an optimized subset of features based on the different samples in every bootstrap, which have a tendency to generate different results with respect to rank aggregation. With discrete wavelet transform, the experiments based on binary and multi-class tasks show that the AHFSE performs well on the Bonn data set and improves the specificity, sensitivity, or accuracy of the selected features by combining the subsets of different feature selections to obtain new samples within the bagging process. Furthermore, the adaptive process helps the framework obtain the optimum combination of the feature selection algorithm. The AHFSE also obtains more desirable final results in several perspectives, such as: 1) compared with other feature selection methods; 2) compared with other ensemble methods; and 3) compared with other research that uses discrete wavelet transform as a preprocessing step.

INDEX TERMS Epileptic seizure detection and classification, discrete wavelet transform, hybrid feature selection, classifier ensemble, bagging, rank aggregation, adaptive, genetic algorithm, optimization, machine learning.

I. INTRODUCTION

Fisher *et al.* [1] explained epilepsy as a neurological disorder group that can be characterized by epileptic seizures in consequence of spontaneous with peculiar cortical nerve cell activities in the brain. The electroencephalogram (EEG) signal is widely used in epileptic seizure detection, as epilepsy diagnosis can be performed by identifying the abnormalities of EEG signals [2]. As manual detection by an expert neurologist is expensive and time consuming and as there are concerns about a possible loss of accuracy caused by fatigue, computer-aided approaches to epilepsy diagnosis using EEG signal time series data analysis can serve as effective solutions to the above problems. Epileptic seizure EEG signal datasets that are widely used to develop seizure detection approaches

mostly contain several groups of subjects, such as: 1) epileptic subjects during seizure-free intervals, 2) epileptic subjects during seizure (ictal) and 3) healthy subjects.

The epileptic seizure research mainly divide into two directions: 1) focuses on preprocessing methods and 2) focus on the classification & detection framework. In the first direction, many researchers introduces the effective approaches to extract the features from raw EEG signal, for example: Bhattacharyya *et al.* [3] computes complexity of the signal which obtained from tunable-Q wavelet transform in multiple oscillatory scales; Bhati *et al.* [4] applied three-band wavelet filter banks which localized from wavelet to alleviate the poor frequency resolution in the high and low frequencies; Sharma & Pachori [5] focuses on feature extraction from the phase space reconstruction (PSR) of intrinsic mode functions; Samiee *et al.* [6] proposed a signal models which based on rational orthogonal polynomials and applied short-time Fourier transform which relied on rational functions; Bhati *et al.* [7] proposed equivalent summation based on uncertainty principle of discrete-time, then compute the frequency variance from discrete Fourier transform. In the second direction, many researchers introduces the robust approaches to classify and detect the epileptic seizure, such as: Abdulhay *et al.* [8] applied stacking-correspondence analysis and nearest neighbor (SCANN) with KNN, Naive Bayes and SVM as base classifiers; Li *et al.* [9] applied modified artificial neural network into bagging ensemble to prevent the detection algorithm from local minima; Peker *et al.* [10] applied complex-valued artificial neural networks due to complex number are important characteristics in the signal processing.

From the description above, this research belong to second direction, which is the purpose of this research is to identify a framework and algorithm with better performance in matter of accuracy, sensitivity and specificity than other methods. In this paper, we focus on classifier ensemble learning with feature selection. Although most researchers have only considered predefined features, we combine ensemble learning with feature selection to improve the performance of epileptic seizure detection. We make two contributions. First, combining the subsets of different feature selections to obtain new samples within the bagging process improves the specificity, sensitivity or accuracy of the selected features. Second, the adaptive process helps the framework obtain the optimum combination of the feature selection algorithm.

Specifically, we design a new algorithm, referred to as the Adaptive Hybrid Feature Selection based classifier Ensemble (AHFSE), for epileptic seizure classification which described at Figure 1. First, to decompose the raw EEG signals into several sub-bands of signals, we applied discrete wavelet transform (DWT). Then, several features were extracted from sub-bands and put into EEG training data. Second, Bagging method were adopted to split the training data into multiple training set (bootstrap) as sample subset. Third, Adaptive hybrid feature selection based on Rank aggregation were applied within every sample subset to generate compact sample subsets (which contains selected feature subsets), then the compact sample subsets are fed into base classifier to obtain the learning model. Fourth, EEG testing data (which already processed using DWT and contains features same in the first step) are sent into every bootstrap, the features from EEG testing data are adjusted according feature subsets in every bootstraps, then fed into predictor (base classifier) to obtain prediction. Finally, voting is then used for final prediction to complete the seizure classification and detection tasks. Several measurement tests are also adopted to compare our framework to other approaches.

From the experiments, our framework obtained high performance when: 1) compared with the feature selection problem; 2) compared with other ensemble methods and 3) with other seizure detection research using discrete wavelet transform as preprocessing methods. Moreover, our framework is comparable with other researchers which used other preprocessing methods.

The remainder of this paper is organized as follows. Section II briefly reviews EEG preprocessing and classification. Section III describes the detailed methods which utilized in this research. Section IV provides the conducted experiment in this research. Section V presents our conclusion and proposes future works.

II. RELATED WORKS

Epileptic seizure classification and detection research is mainly divided into three steps: preprocessing, feature extraction and classification for detection and prediction. The first step is preprocessing, which is used to analyze several changes in EEG signals. The second step is feature extraction, which is used to obtain important features that can maximize the separability of different classes. Mostly, feature extraction is described within preprocessing methods, one of those categories is time-frequency (wavelet) domain methods, which decompose EEG signals into sub-band signals with down sampling. Discrete wavelet transform (DWT) are widely used for EEG signal analysis and belong to the class of wavelet domain methods [11]. Statistical features [9], [11] and nonlinear dynamics, such as entropy [12], [13], fractal dimension [8], [14], are commonly applied in the feature extraction process. The experimental outcomes of these references show that DWT with those features can accurately characterize epileptic seizure signals.

As a good subset of features can improve the accuracy of epileptic seizure detection, some researchers have used feature selection to identify feature subsets that can improve accuracy. For example, Alzami *et al.* [15] adopted a modified differential evolution feature selection approach to obtain the optimum feature subset for improving the robustness of epileptic seizure detection and classification accuracy. Lee *et al.* [16] applied a neural network with a fuzzy membership function to improve classification performance based on feature selection. Pippa *et al.* [17] used the Relief-F algorithm to estimate the importance of each feature in epileptic seizure classification. Guo *et al.* [18] used genetic

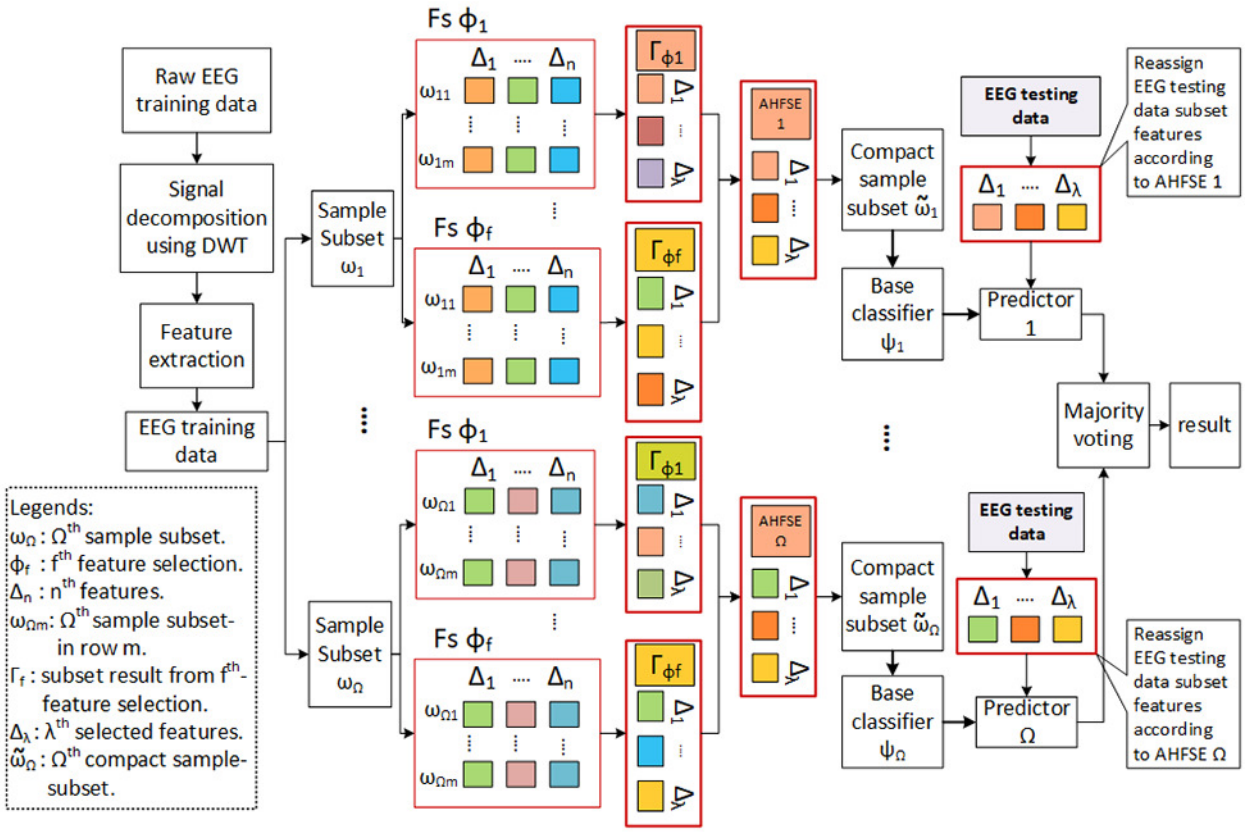


FIGURE 1. Block diagram of adaptive hybrid feature selection-based ensemble.

programming to select features for reducing the dimensions number.

The third step is classification for detection and prediction, which is used to classify EEG signals and determine whether they are indicative of a seizure. The classification method is divided into two categories: 1) single classifiers and 2) ensemble learning classifiers.

The single classifier method mainly focuses on which features can obtain the highest accuracy and which optimum parameters should be applied to classifiers. For example, Guler and Ubeyli [19] used multi-class SVM and the probabilistic neural network to classify seizures with the wavelet coefficient and Lyapunov exponent as the features. Gosh-Dastidar *et al.* [20] applied principal component analysis to enhance and transform nine features, and then fed them into the cosine radial basis function (RBF) neural network to detect seizure.

Ensemble learning method is used to unify multiple classifiers, which provides more robust, stable and accurate final results than a single classifier. Ensemble learning approaches have been applied in many areas, particularly in bioinformatics and health informatics [21], [22]. In epileptic seizure detection and classification, Hassan *et al.* [23] applied bagging combined with a tunable-Q factor wavelet transform to detect epileptic seizures, Abdulhay *et al.* [8] used bagging in which the prediction of a KNN-SVM-NB module is fed into a meta learning module to obtain the final detection result and Li *et al.* [9] adopted a neural network ensemble that prevents the detection algorithm from falling into a local minimum.

This research proposes a new approach in the epileptic seizure classification which is dissimilar from the methods which presented above. In this research, Adaptive hybrid feature selection-based ensemble are used to classify the seizure. The proposed framework produced better performance comparing other feature selection (can be seen at Table 3), other ensemble methods (Table 5) and other research which using discrete wavelet transform as preprocessing step (Table 7). Following description is a summary of proposed framework.

This research explores the effects of combining feature selection using rank aggregation which chosen adaptively. The proposed framework adopt a rank-aggregation (RA) methods from Kolde [24]. Most of filter feature selection uses different assumptions to obtain the best subset of features. Thus, combining their results using RA would cover a larger set of possible underlying data, thereby improving the specificity, sensitivity or accuracy of the selected features. In our research, we applied the RA within bagging process, then, genetic algorithm is applied in search of best combination of feature selection to calculate the reliability of the subset results and applicability to partial ranking. This genetic

algorithm together with RA is described at sections III-C. Sections III-D describes the advantages and limitations of this framework including time complexity analysis. In here, we applied discrete wavelet transform (DWT) as preprocessing methods, thus in section IV-B5 we will describes the advantages and disadvantages the DWT with other preprocessing methods.

III. METHODS

A. DISCRETE WAVELET TRANSFORM

Wavelet transforms (WT) is a good feature extractor, at low frequencies, WT able to gives proper frequency information, and at high frequencies, WT gives proper time information. Thus, WT is compatible for analysis non-stationary data patterns. Discrete wavelet transform (DWT) is computationally cheap rather than continuous wavelet transform (CWT) because DWT applied shifts and scales is selected based on power of two.

$$DWT(S_c, S_f) = \frac{1}{\sqrt{|S_c|}} \int_{-\infty}^{\infty} x(t)\psi_f \left(\frac{t - S_f S_c}{S_c}\right) dt \quad (1)$$

In here, S_c is scaling parameters, S_f is shifting parameters, $x(t)$ is a signal and ψ_f is mother wavelet. Then, wavelet coefficients can be obtained using Daubbechies family. Thus, the obtained wavelet coefficient is used to extract the desired features.

B. FEATURE SELECTIONS

In this subsection, we will briefly described several feature selection that widely used, such as: Minimal redundancy maximum relevance (mRMR) feature selection [25] uses mutual information to calculate the dependence between features and labels. The features with lowest redundancy (mutually maximally dissimilar) and highest relevance are assigned as selected features. We can calculate the redundancy and relevance as follows:

$$W_I(\Gamma) = \arg\min\{\frac{1}{|\Gamma|^2} \sum_{a,b\in\Gamma} I(a,b)\} \quad (2)$$

$$V_I(\Gamma, h) = \arg\max\{\frac{1}{|\Gamma|} \sum_{a\in\Gamma} I(h,a)\} \quad (3)$$

where Γ is the set of features we are seeking, W_I denotes redundancy, $I(a, b)$ is the mutual information between features a and b, V_I denotes relevance and h represents the target classes. The mRMR feature set can be obtained by simultaneously optimizing Eq. (2) and Eq. (3) as follows:

$$\vartheta = \max_{a\in S_\Gamma} [I(a,h) - \frac{1}{|\Gamma|} \sum_{b\in\Gamma} I(a,b)] \quad (4)$$

where ϑ represents the near-optimal features and S_Γ represents the set features minus one feature (as we already obtain one feature from the previous calculation).

Fisher [40] used Fisher score performance criteria to select individual features. The Fisher score provides a feature subset, in which distance is minimized for features that belong to the same class. The selected features must maximize the Fisher score of the overall subset. We can calculate the Fisher score of the Δ-th feature, S_Δ, as follows:

$$S_\Delta = \frac{\sum_{k=1}^{K} d_y(\mu_{\Delta y} - \mu_\Delta)^2}{\sum_{k=1}^{K} d_y \varrho_{\Delta y}^2} \quad (5)$$

where d_y is the number of individuals in the y-th class, K is the number of classes, $\mu_{\Delta y}$ and $\varrho_{\Delta y}$ are the mean and variance of the Δ-th feature in the y-th class, and $\mu_{x\Delta}$ is the mean of the Δ-th feature. As the Fisher score is a univariate scheme, Fisher feature selection cannot handle feature redundancy.

Relief-F feature selection [26] randomly chooses individuals and then finds the nearest neighbor individual in the same class and the non-similar nearest neighbor individual from a different class. Relief-F then updates the weights of every individual. Finally, individuals that obtain higher relevance (mean of weights) are chosen as selected features. We can calculate the Relief-F score of the Δ-th feature, S_Δ, as follows:

$$S_\Delta = \frac{1}{K} \sum_{k=1}^{r} (S_{\Delta_a} + S_{\Delta_b}) \quad (6)$$

$$S_{\Delta_a} = (-\frac{1}{m_k} \sum_{X_y\in M_k} \lambda(\Delta_{\Delta k} - \Delta_{\Delta y})) \quad (7)$$

$$S_{\Delta_b} = \sum_{Y\neq Y_k} \frac{1}{h_{kY}} \frac{p(Y)}{1 - p(Y)} \sum_{\Delta_y\in H_k} \lambda(\Delta_{\Delta k} - \Delta_{\Delta y}) \quad (8)$$

where r is the individual that is randomly sampled, λ is the number of features to be selected, M_k and h_{kY} are the set of imminent points to Δ_k with the same class and class Y with the sizes of m_k and h_{kY}, respectively, and $p(Y)$ is the possibility of individuals being from class Y.

Chi-square feature selection [27] is a numerical test that evaluate deviation from the prospective distribution, considering that the feature event is independent of the class value. If the two events are dependent, it should be helpful as a feature. High chi-square values represent the increased likelihood of the features being correlated with the class. The features are sorted in decreasing order of chi-square value to compare the importance of the features. We can calculate the chi-square, χ^2, score of the Δ-th feature, S_Δ, as follows:

$$S_\Delta = \frac{d(\varphi d - \upsilon P)^2}{P\upsilon(d - P)(d - \upsilon)} \quad (9)$$

where d is the total number of individuals, φ is the total number of positive individuals demonstrating feature Δ, υ is the total number of individuals demonstrating feature Δ and P is the total number of positive instances.

As remainder, S_Δ is a score of features, not assigned features. Thus, we can sorted the features according to the scores and put selected features as chosen subset features Γ.

C. ADAPTIVE HYBRID FEATURE SELECTION-BASED ENSEMBLE

Algorithm 1 provides an overview of our proposed framework of seizure classification. This framework divided into five steps: preprocessing, obtaining sample subsets ω, obtaining compact sample subset $\widetilde{\omega}$, learning model and classification (prediction).

In the preprocessing step, Raw EEG training data D_{tr} and raw EEG testing data (for classification/prediction) D_{ts} are decomposed into sub-band signals, then extract features from decomposed D_{tr} and D_{ts} and put into training data $D_{tr}(\Delta_a)$ and testing data $D_{ts}(\Delta_a)$ (where $a \in \{1, \ldots,$ number of extracted features $\}$). Then $D_{tr}(\Delta_a)$ and $D_{ts}(\Delta_a)$ is normalized using scale σ.

In the obtaining sample subsets ω step, bagging method is applied to training data $D_{tr}(\Delta_a)$ to generate sample subsets ω_i (where $i \in \{1, \ldots, \Omega\}$ and Ω is the number of new bootstraps (new sample subsets)).

After the new sample subsets is generated, the next step is obtaining compact sample subset $\widetilde{\omega}_i$. In here, we need call Algorithm 3 (Adaptive Hybrid Feature Selection based Ensemble/AHFSE). The AHFSE will produce compact sample subset $\widetilde{\omega}_i$ which utilized rank-aggregation together with genetic algorithm. It means, in every $\widetilde{\omega}_i$ could have different feature subsets. For example, suppose we need to obtain 6 number of subsets λ, the numbers of features are 45 and number of bootstraps Ω are 50. Given different sample combination in every sample subset ω_i, the AHFSE high likely will produce different feature subsets Γ according to ω_i, for example: $\omega_{1(\Gamma)} = \{23, 5, 14, 32, 8, 41\}$; $\omega_{2(\Gamma)} = \{5, 14, 23, 32, 8, 45\}$; . . .; and $\omega_{50(\Gamma)} = \{14, 5, 23, 32, 41, 8\}$.

When all compact sample subset $\widetilde{\omega}_i$ is generated, $\widetilde{\omega}_i$ will be fed into training model (base-classifier) Ψ_i (where $i \in \{1, \ldots, \Omega\}$) to obtain the learning model h_i.

In the classification (prediction) step, the $D_{ts}(\Delta_a)$ is fed into every predictor (from $1, \ldots, \Omega$). In here, feature subsets of $D_{ts}(\Delta_a)$ is adjusted according the $\widetilde{\omega}_i$ in respective matter. It means, $D_{ts}(\Delta_a)$ in predictor 1 will have feature subsets same like $\omega_{1(\Gamma)}$, which can be seen as $D_{ts(1)(\Gamma)} = \{23, 5, 14, 32, 8, 41\}$. In other words, $D_{ts(i)}$ is the testing dataset which placed in the predictor i. Then, label prediction for $D_{ts(i)}$ is generated respected with h_i. Finally, majority voting is used to obtain the final results.

Algorithm 3 provides an overview of AHFSE, which used to obtain compact sample subset $\widetilde{\omega}$. First, we map feature selection $\Phi = \{\phi_1, \phi_2, \ldots, \phi_f\}$ onto individuals in population α_Λ as follows:

$$\alpha_\Lambda = \{Z_\phi^1, Z_\phi^2, \ldots, Z_\phi^f\} \tag{10}$$

where Z_ϕ^1 means that ϕ_1 is placed in the first binary vector, Z_ϕ^2 means that ϕ_2 is placed in the second binary vector and so on. Then, the initial population, pop_Λ, is generated with indicator Z^f, where 1 denotes that the binary vector (feature selection) is selected and 0 denotes that it is not selected. Second, until terminal condition η_Λ is satisfied, we must generate the crossover individual pop_γ and mutation individual pop_μ, put all into $pop_\Lambda = [pop_\Lambda, pop_\gamma, pop_\mu]$, sort pop_Λ and reduce the population of pop_Λ corresponding to α_Λ. We must then define the first individual as the best individual, α_{best}. Third, after η_Λ is satisfied, the final $\alpha_{best} = \{Z_\phi^1, Z_\phi^2, \ldots, Z_\phi^l\}$ (where l is the position of the binary vector that is selected) is transformed into the selected feature selection algorithm as follows:

$$\Phi_\Lambda = \{\phi_1, \phi_2, \ldots, \phi_s\} \tag{11}$$

where $\phi_1 = Z_\phi^1$, $\phi_2 = Z_\phi^2$, and $\phi_s = Z_\phi^l$. In other words, Φ_Λ is collection of selected feature selection that generated from AHFSE.

After we obtained Φ_Λ, next step is applied rank-aggregation as follows: First, ω_i is fed into every feature

Algorithm 1 Proposed Framework of Seizure Classification

Require:
Input: Raw EEG training data, D_{tr}
Raw EEG testing data, D_{ts}
the scaling options for the normalized dataset, σ;
the number of bootstraps, Ω;
the set of feature selections, $\Phi = \{\phi_1, \phi_2, \ldots, \phi_f\}$;
the number of subsets to obtain, λ;
the set of weak classifiers, $\Psi = \{\psi_1, \psi_2, \ldots, \psi_\Omega\}$;

Ensure:
/* step: preprocessing */
decompose D_{tr} & D_{ts} using Eq.1;
extract features of decomposed D_{tr} and put into EEG training data $D_{tr}(\Delta_a)$; /*(where $a \in \{1, \ldots,$ number of features$\}$)*/
extract features of decomposed D_{ts} and put into EEG testing data $D_{ts}(\Delta_a)$;
whether $D_{tr}(\Delta_a)$ and $D_{ts}(\Delta_a)$ is already processed depends on σ;
/* step: obtaining sample subset ω */
generate Ω bootstrap sampling from $D_{tr}(\Delta_a)$ as ω_i; /*(where $i \in \{1, 2, \ldots, \Omega\}$)*/
/* step: obtaining compact sample subset $\widetilde{\omega}$ */
For ω_i in $1, \ldots, \Omega$
    call algorithm 3 to obtain compact sample subset $\widetilde{\omega}_i$;
EndFor Ω
/* step: learning model */
For $\widetilde{\omega}_i$ in $1, \ldots, \Omega$
    train base classifier ψ_i using $\widetilde{\omega}_i$ to obtain h_i;
EndFor Ω
/* step: classification (prediction) */
For $\widetilde{\omega}_i$ in $1, \ldots, \Omega$
    reassign feature subsets $D_{ts}(\Delta_a)$ according to $\omega_{i(\Gamma)}$ as $D_{ts(i)}$; /*(where $\omega_{i(\Gamma)}$ is feature subset result - from AHFSE)*/
    label prediction using h_i and $D_{ts(i)}$;
EndFor Ω
apply a majority voting scheme to obtain final results;
Output: final prediction results;

selection algorithm that provided from Φ_Λ corresponding with number of subset of features λ to obtain the new subset features, which can be seen as:

$$\Gamma_i = \{\Delta_{b,1}, \Delta_{b,2}, \ldots, \Delta_{b,\lambda}\} \tag{12}$$

(where $b \in \{1, \ldots, s\}$ and s is the number of selected feature selection). As many subset features in Γ_i are redundant and selected in different assumptions, we must purify the subset features as $\mu_{[\tau \times s]}$ by using unique elements, such as γ and map number δ of Γ_i. Here $\Gamma_{i_{pos}}$ is the number of feature subsets obtained from Γ_i. γ returns the same data as Γ_i but with no repetitions. τ is the number of elements. γ and δ (where $\delta \in 1, 2, \ldots \Gamma_{i_{pos}}$) are indexed to γ, which returns as a column vector. Then δ is transformed into the μ matrix as follows:

$$\mu_{[\tau \times s]} = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1s} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2s} \\ \vdots & \vdots & \vdots & \ddots \\ \delta_{\tau 1} & \cdots & \cdots & \delta_{\tau s} \end{bmatrix} \tag{13}$$

After the feature subsets is purified and transformed into $\mu_{[\tau \times s]}$, we must obtain the importance (rank) matrix $\Theta_{[\tau \times s]}$ by using the unique elements, U, and map number, T, of μ, where U returns the same data as μ but with no repetitions. U_l is the length of U and T is indexed to U, which returns as column vector. Then, T is reshaped into (τ by s) matrix as ($\Xi_{[\tau \times s]}$). As matrix $\Xi_{[\tau \times s]}$ only contains the map number, T (where T ranges from $\{1, 2, \ldots, U_l\}$), we must map matrix $\Xi_{[\tau \times s]}$ to matrix $\xi_{[U_l \times s]}$ with respect to $V = \{1, 2, \ldots, \tau\}$ (where V is the position of γ) as follows:

$$\xi_{[\tau \times s]} = \sum_{j=1}^{s} \sum_{i=1}^{\tau} \{\xi([\Xi_{i,j}, j])\} \tag{14}$$

$$[\Xi_{i,j}, j] = V_i \tag{15}$$

As Eq. (14) fills matrix ξ as much as $\xi_{[\tau \times s]}$, the empty matrix position of ξ can be filled with τ. Finally, all $\xi_{[U_l \times s]}$ values are obtained. Here, $\xi_{[U_l \times s]}$ is the priority rank of γ, meaning that we can reduce the ξ matrix size from $[U_l \times s]$ to $[\tau \times s]$. The final matrix of $\xi_{[\tau \times f]}$ can be defined as follows:

$$\xi_{[\tau \times s]} = \begin{bmatrix} \Xi_{11} & \Xi_{12} & \cdots & \Xi_{1s} \\ \Xi_{21} & \Xi_{22} & \cdots & \Xi_{2s} \\ \vdots & \vdots & \vdots & \ddots \\ \Xi_{\tau 1} & \cdots & \cdots & \Xi_{\tau s} \end{bmatrix} \tag{16}$$

Finally, we can obtain the rank matrix as follows:

$$\Theta_{[\tau \times s]} = \arg \sum_{i=1}^{\tau} \sum_{j=1}^{s} \{\frac{\Xi_{i,j}}{U_l}\} \tag{17}$$

This then becomes the following:

$$\Theta_{[\tau \times s]} = \begin{bmatrix} \Theta_{11} & \Theta_{12} & \cdots & \Theta_{1s} \\ \Theta_{21} & \Theta_{22} & \cdots & \Theta_{2s} \\ \vdots & \vdots & \vdots & \ddots \\ \Theta_{\tau 1} & \cdots & \cdots & \Theta_{\tau s} \end{bmatrix} \tag{18}$$

Here, we can obtain subset features after we execute Eq. (14) as pre-ranked feature subsets $\Gamma_{i_{pre}}$ which have following values:

$$\Gamma_{i_{pre}} = \{\gamma_1, \gamma_2, \ldots, \gamma_\tau\} \tag{19}$$

Because γ is actually unique elements from Γ_i, we can transform Eq. (19) into:

$$\Gamma_{i_{pre}} = \{\Delta_1, \Delta_2, \ldots, \Delta_\tau\} \tag{20}$$

Furthermore, we must obtain the statistical significance (ρ value) of the mean rank, $\Theta_{[\tau \times s]}$, by calculating the ρ scores. The ρ scores can be obtained as described as Algorithm 2.

Algorithm 2 Calculate ρ Scores

Require:

Input: ranking matrix $\Theta_{\tau \times s}$;

Ensure:

1: **For every row of** $\Theta_{\tau \times s}$:
2: Calculate the beta scores of the current row using the beta cumulative distribution (beta-CDF) function to obtain η;
3: Find the minimum value of η, η_{min};
4: Determine how many η variables are not NaN, ϖ;
5: obtain the ρ values using the beta-CDF function with η_{min} as the value and parameters 1 and ϖ, respectively;
6: **EndFor** row
7: $\widetilde{\rho} = \{\rho_1, \rho_2, \ldots, \rho_\tau\}$;

Output: $\widetilde{\rho}$

After we obtain the statistical significance, we need put $\widetilde{\rho}$ together with $\Gamma_{i_{pre}}$ as $\Upsilon = [\widetilde{\rho}, \Gamma_{i_{pre}}]$. The Υ can be seen as Δ_1 have the statistical significance value as ρ_1, and Δ_τ have the statistical significance value as ρ_τ. Then, we need to sort the $\widetilde{\rho}$ in ascending order as the lower number corresponds to a better ranking. As we only need the λ subsets, the final subset features can be shown as follows:

$$\Gamma_{i_{(AHFSE)}} \in \{\Delta_1, \Delta_2, \ldots, \Delta_\lambda\} \tag{21}$$

Finally, compact sample subsets $\widetilde{\omega_i}$ can be obtained by reassign the features subsets $\Gamma_{i_{(AHFSE)}}$ into ω_i, which can be seen as:

$$\widetilde{\omega_i} = (\omega_i, \Gamma_{i_{(AHFSE)}}) \tag{22}$$

Algorithm 4 provides an overview of obtaining fitness cost. First, we must map the information from the current α_Λ to the individual α_{indv} by determining which binary vector is selected. The list of selected feature selections, ϕ_G, can then be obtained as follows:

$$\phi_G = \{\phi_{g(1)}, \phi_{g(2)}, \ldots, \phi_{g(\alpha_{indvSel})}\} \tag{23}$$

where $\alpha_{indvSel}$ is the number of selected binary vectors. Then, we obtain subsets from every feature selection, ϕ_g, resulting in:

$$\Gamma_{\phi_{g(j)}} = \{\gamma_{j,1}, \gamma_{j,2}, \ldots, \gamma_{j,\lambda}\} \tag{24}$$

Algorithm 3 Adaptive Hybrid Feature Selection Based Ensemble (AHFSE)

Require:
Input: sample subset, ω_i;
the set of feature selections, $\Phi = \{\phi_1, \phi_2, \ldots, \phi_f\}$;
the number of subsets to obtain, λ;
the individual length, f;
the number of iterations, η_Λ;
the size of population, α_Λ;
the rate probability of crossover, γ_Λ;
the rate probability of mutation, μ_Λ;

Ensure:
1: map the feature selection, Φ, onto individuals in population α_Λ as Eq. (10);
2: Initialization step:
3: generate α_Λ population randomly;
4: evaluate each individual fitness, θ, using Algorithm 4;
5: save the individual and θ to population pop_Λ;
6: sort pop_Λ based on θ;
7: obtain best solution α_{best} from first individual;
8: **For** i=1 to η_Λ:
9: obtain pop_γ by using the crossover scheme with γ_Λ;
10: obtain pop_μ by using the mutation scheme with μ_Λ;
11: merge the population, $pop_\Lambda = [pop_\Lambda, pop_\gamma, pop_\mu]$;
12: sort population pop_Λ based on θ;
13: reduce population, pop_Λ, corresponding to α_Λ;
14: define the first individual as the best individual α_{best};
15: **EndFor** η_Λ /* termination condition achieved */
16: obtain α_{best};
17: transform α_{best} into selected Φ_Λ following Eq. (11)
Output: selected features $\Gamma_i = \{\Delta_1, \ldots, \Delta_\lambda\}$

where $j \in \{1, \ldots, \alpha_{indvSel}\}$. After we obtain subsets from every feature selection $\Gamma_{\phi_{g(j)}}$, we need to put all $\Gamma_{\phi_{g(j)}}$ into one bucket as follows:

$$\Gamma_{\phi_{(G)}} = \{\Gamma_{1,1}, \Gamma_{1,2}, \ldots, \Gamma_{1,\lambda}, \Gamma_{2,1}, \ldots, \Gamma_{2,\lambda}, \ldots, \Gamma_{f,\lambda}\} \tag{25}$$

The next step is to obtain the new subset features, $\widetilde{\Gamma}_{i(G)}$, of $\Gamma_{\phi_{(G)}}$ using Eq. (13) - (21). We then remap the current ω_i with subset $\widetilde{\Gamma}_{i(G)}$ which can be seen as:

$$\omega_{i(G)} = (\omega_i, \widetilde{\Gamma}_{i(G)}) \tag{26}$$

In here, Eq. (26) can be seen as $\omega_{i(G)}$ consist of sample ω_i with features $\widetilde{\Gamma}_{i(G)}$.

Finally, we feed $\omega_{i(G)}$ as a sample into the classifier and obtain the error result as the fitness cost, θ.

As worth mention, we define our own rule regarding the fitness functions, which is: the genetic algorithm will probably generate empty individual (which means, no feature selection is selected), thus, we gave fitness cost directly to maximum value. As error values is the fitness value, in here we gave result 100% error value (% is considered when we using percentage as accuracy measurement).

Algorithm 4 Fitness Function to Obtain Cost

Require:
Input: current α_Λ individual, α_{indv};
current bootstrap sampling, ω_i;
the number of subsets to obtain, λ;
the set of feature selections, $\Phi = \{\phi_1, \phi_2, \ldots, \phi_f\}$;

Ensure:
1: map the information from α_{indv} to become ϕ_G following Eq. (23)
2: obtain the subset features following Eq. (24)-(25)
3: obtain the new subset features, $\widetilde{\Gamma}_{i(G)}$, using Eq. (13) - (21).
4: remap the current ω_i to $\omega_{i(G)}$ using Eq. (26)
5: feed $\omega_{i(G)}$ into the classifier and obtain the error value as fitness cost, θ
Output: fitness cost, θ

TABLE 1. Experiments parameter settings.

Parameters	Values
Number of feature subsets when using feature selection, λ	6 & 10
Number of feature subsets when not using feature selection	45
Number of bootstraps, Ω	50

TABLE 2. AHFSE parameters settings.

Parameters	Values
Length of individual, $[1f]$	4
Iteration	10
Population number $\leftarrow ((f * 2) - 1)$	7
Crossover probability	0.7
Mutation probability	0.1

D. COMPLEXITY ANALYSIS

We also performed a complexity analysis of our framework with respect to its computational cost. The corresponding time complexity is computed as follows:

$$T_{framework} = T_{WT} + T_{OS} + T_{OCS} + T_{LM} + T_{CP} \tag{27}$$

The explanation as follows: (1) T_{WT} denote the computational cost of converting raw EEG signal into dataset which contains sample and features using wavelet transform. T_{WT} is related to the number channel of EEG CL, sub-band signal SB. In here, for every number of channel, we need to decomposed those raw EEG signal into sub-band signals. Then in every sub-band signal, we need to extract the features Δ_b (where $b \in \{1, 2, \ldots$ end of features$\}$).

$$T_{WT} = O(CL.(SB.(\Delta_b))) \tag{28}$$

(2) T_{OS} denote the computational cost of generating new sample subsets. T_{OS} is related to the number of bootstraps Ω, which is as follows:

$$T_{OS} = O(\Omega) \tag{29}$$

(3) T_{OCS} denote the computational cost of generating new compact sample subsets. T_{OCS} is related to the number of

TABLE 3. Comparison with other feature selections.

classification task		None	AHFSE		Chi-Square		Fisher		mrmr		Relief-F	
task type	measure	45	6	10	6	10	6	10	6	10	6	10
E-A	Acc	97.50(0.41)	**100**	**100**	99.33(0.47)	99.17(0.24)	99.33(0.47)	99.67(0.47)	99.17(0.24)	98.83(0.47)	99.5(0.41)	99.33(0.24)
	Sen	99.29	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	Spec	99.59	**100**	**100**	98.73	98.41	98.69	99.33	98.45	97.83	99.02	98.61
E-B	Acc	97.5	**99.5(0.41)**	**99.5**	99.00(0.41)	98.50(0.71)	99.33(0.24)	99.00(0.41)	98.67(0.24)	98.67(0.24)	99.33(0.24)	99.33(0.24)
	Sen	**100**	**100**	**100**	99.65	**100**	**100**	99.65	99.3	**100**	**100**	**100**
	Spec	94.99	**99.01**	**99.01**	95.87	96.94	98.72	98.38	98.25	97.31	98.72	98.68
E-C	Acc	96.67(0.47)	98	**98.5**	98.33(0.24)	98.17(0.47)	97.67(0.47)	97.83(0.62)	98(0.41)	98(0.41)	98(0.41)	97.83(0.62)
	Sen	97.7	97.76	98.83	**98.96**	98.63	97.14	98.44	98.38	98.58	98.1	98.44
	Spec	95.5	**98.11**	98.1	97.83	97.63	98.01	97.15	97.68	97.5	97.8	97.25
E-D	Acc	95.17(0.62)	**97.33(0.24)**	96.17(0.62)	97.00(0.71)	96.17(0.24)	96.83(0.47)	95.67(0.24)	96	95.17(0.10)	**97.33(0.24)**	95.67(0.62)
	Sen	96.78	96.76	96.3	96.74	96.36	97.29	96.2	95.58	95.52	96.9	96.06
	Spec	93.56	97.49	96.07	96.87	95.67	96.53	95.21	96.16	94.48	97.5	95.45
E-CD	Acc	96.67(0.27)	97.33(0.47)	**98.44(0.16)**	97.56(0.42)	97.22(0.57)	97.33(0.54)	97(0.27)	96.22(0.42)	96.11(0.57)	97.33(0.47)	97.56(0.83)
	Sen	97.14	95.32	**97.46**	96.23	96.95	95.68	95.8	94.73	95.08	95.32	95.19
	Spec	96.55	98.3	**98.99**	98.34	97.52	98.15	97.46	97	96.65	98.3	98.66
E-ABCD	Acc	98.13(0.09)	98.6(0.43)	**99.00(0.06)**	97.87(0.009)	97.67(0.19)	98.8(0.16)	98.53(0.09)	97.73(0.19)	97.27(0.09)	98.8(0.16)	98.93(0.19)
	Sen	97.38	96.67	97.32	94.46	94.95	96.93	97.3	94.02	95.02	**97.4**	96.35
	Spec	98.35	99.09	99.26	98.76	98.35	99.25	98.85	98.75	97.86	99.19	**99.59**
E-A-D	Acc	87.33(0.27)	96.00(0.13)	93.67(0.27)	**97.00(0.27)**	95.00(0.14)	95.22(0.57)	92.00(0.11)	94.33(0.27)	93.56(0.12)	95.44(0.42)	93.11(0.42)
	Sen	**98.76**	98.27	97.27	97.23	97.6	97.88	96.51	95.64	95.62	97.58	97.56
	Spec	97.41	**98.87**	98	98.84	98.26	98.53	97.5	98.39	98.24	98.33	97.68
E-AB-CD	Acc	93.27(0.47)	**97(0.28)**	96.67(0.19)	96.53(0.19)	96.67(0.57)	97.00(0.53)	96.27(0.66)	92.8(0.33)	92.4(0.43)	96.93(0.09)	96.07(0.38)
	Sen	**99.03**	96.53	97.19	96.13	96.87	97.45	97.36	95.74	92.19	96.64	94.81
	Spec	98.19	98.93	98.94	98.85	98.76	**99.01**	98.6	94.75	94.82	96.89	98.93

bootstraps Ω, AHFSE which containing numbers of population α_Λ, numbers of genetic iteration η_Λ, number of crossover γ_Λ, number of mutation μ_Λ and numbers of iteration of classifier to obtain the fitness cost θ_{run}.

$$T_{OCS} = O(\Omega.(T_{pop} + T_{iter})) \tag{30}$$

$$T_{pop} = O(\alpha_\Lambda.(\theta_{run})) \tag{31}$$

$$T_{iter} = O(\eta_\Lambda.(\gamma_\Lambda.(\theta_{run}) + \mu_\Lambda.(\theta_{run})) \tag{32}$$

(4) T_{LM} denote the computational cost of generating learning model. T_{LM} is related to the number of bootstraps Ω which is follows:

$$T_{LM} = O(\Omega) \tag{33}$$

(5) T_{CP} denote the computational cost of making predictions. T_{CP} is related to the number of bootstraps Ω which is follows:

$$T_{CP} = T_{WT} + O(\Omega) \tag{34}$$

The reason T_{CP} also contains T_{WT} is due to testing dataset is obtained directly from patient. Thus, when the framework is put into real-time monitor device, we only need concern in Eq. (34). It also worth mentions, when considering Eq. (28) - (33), parallel processing could be applied to decrease the computational time.

The advantages of our framework is: 1) successfully obtained the best subset of features in every bootstrap and placed as compact sample subset; and 2) It can be used using many feature selection not limited in those feature selections described in Section III-B. The reason we also explained the several feature selection is to grasp the understanding of filter feature selection. Thus, our framework can be used using other filter feature selection. The limitations of our framework is: 1) need longer time obtain compact sample subsets which is T_{OCS}, where it can be seen at every bootstrap Ω, need to generate number of population α_Λ and genetic iteration η_Λ; 2) in matter of detection, AHFSE slightly need more time rather than simple bagging due to the feature subset of test data need to be reassigned following the AHFSE in current bootstrap location.

TABLE 4. Average rankings of the feature selection (Friedman).

Algorithm	Ranking (subset 6)	Ranking (subset 10)
AHFSE	**2.0625**	**1.25**
Chi-Square	2.8125	2.875
Relief-F	2.4375	3.125
Fisher	3.0625	3.5
mRMR	5.125	4.8125
None	5.5	5.4375

IV. RESULTS & DISCUSSION

A. DATASET

We used a freely available on-line EEG dataset provided by the University of Bonn [28]. The EEG dataset consists of five sets (A-E), each of which contains 100 single-channel EEG segments with a duration of 23.6 s per segment. Each epoch, 4, 096 samples were generated using 173.61 Hz. For the healthy dataset, five healthy volunteers participated with opened eyes (denoted as set A) and closed eyes (denoted as set B). Five epileptic seizure patients were also chosen for presurgical assessment for epilepsy using intracranial electrodes. Then, the electrodes were placed symmetrically to record EEG signals from the hippocampal position of the opposite hemisphere of the brain (denoted as set C) and from the epileptic zone (denoted as set D). Set E was taken from all electrodes. Those in sets C and D demonstrated interictal intervals(seizure-free), whereas seizure activities occurred in set D.

To preprocess the raw EEG signal dataset, the band-pass FIR filter is set from range 0.5 Hz to 60 Hz. DWT *db*4 level 4

TABLE 5. Comparison with other ensemble methods.

Task Type	measure	AHFSE	Bagging	Adaboost	Random Forest	Random Subspace
E-A	Acc	**100**	97.50(0.41)	78.83(0.919)	99.5	96.83(0.24)
	Sen	**100**	99.29	**100**	**100**	**100**
	Spec	**100**	99.59	-	99.03	94.25
E-B	Acc	**99.5**	97.5	95.83(0.448)	99.17(0.85)	96.83(0.62)
	Sen	**100**	**100**	99.28	99.37	**100**
	Spec	**99.01**	94.99	-	98.87	94.44
E-C	Acc	98.5	96.67(0.47)	**98.83(0.24)**	98.5	97.33(0.24)
	Sen	98.83	97.7	99.04	98.1	**100**
	Spec	98.1	95.5	98.75	**98.99**	95.08
E-D	Acc	**97.33(0.24)**	95.17(0.62)	**97.33(0.103)**	97.17(0.47)	94.83(0.62)
	Sen	96.76	96.78	97.72	96.02	**98.82**
	Spec	97.49	93.56	97.21	**98.25**	91.56
E-CD	Acc	**98.44(0.16)**	96.67(0.27)	97.44(0.57)	97.44(0.68)	96
	Sen	97.46	97.14	95.75	95.97	**99.65**
	Spec	**98.99**	96.55	98.3	98.37	94.47
E-ABCD	Acc	**99.00(0.06)**	98.13(0.09)	98.6	98.80(0.16)	96.4(0.28)
	Sen	97.32	97.38	96.79	96.86	**100**
	Spec	99.26	98.35	99.09	**99.34**	95.74
E-A-D	Acc	**96.00(0.13)**	87.33(0.27)	92.67(0.47)	95.67(0.54)	81.67(0.98)
	Sen	98.27	**98.76**	95.47	97.31	98.66
	Spec	98.87	97.41	98.97	**99.11**	96.15
E-AB-CD	Acc	**97(0.28)**	93.27(0.47)	93.20(0.28)	96.07(0.52)	87.33(0.93)
	Sen	96.53	99.03	94.35	96.94	**100**
	Spec	98.93	98.19	97.88	**99.34**	97.18

was then used to decompose the filtered EEG signals into five sub-band signals, namely delta, theta, alpha, beta and gamma. Nine features were extracted from every sub-band signal, which contained: kurtosis, skewness, median, mean, standard deviation, Sample Entropy, Embedding Dimension, Correlation Dimensions and Fractal Dimensions.

To evaluate the performance of our framework and comparative researches, we performed different classification tasks. We particularly investigated the classification of seizure epochs (set E) in the presence of other sets. Thus, the classification task is as follows:

- *(E-A)* : set E (seizure) prediction within presence of set A (healthy eyes open) (binary class problem)
- *(E-B)* : set E (seizure) prediction within presence of set B (healthy eyes closed) (binary class problem)
- *(E-C)* : set E (seizure) prediction within presence of set C (seizure-free) (binary class problem)
- *(E-D)* : set E (seizure) prediction within presence of set D (seizure-free) (binary class problem)
- *(E-CD)* : set E (seizure) prediction within presence of set CD (seizure-free) (binary class problem)
- *(E-ABCD)* : set E (seizure) prediction within presence of set ABCD (healthy) (binary class problem)
- *(E-A-D)* : set E (seizure) prediction within presence of set A (healthy eyes open) and set D (seizure-free) (multi-class problem)
- *(E-AB-CD)* : set E (seizure) prediction within presence of set AB (healthy) and set CD (seizure-free) (multi-class problem)

The reason we include the multi-class problems because it is represent the real-world problem. People who suffering from epilepsy can lived normally, and occasionally gotten seizure. Thus by applying the multi-class problem, our long-goal is implemented the framework into device which real-time monitor the people who suffering from epilepsy and could giving immediate response and treatment to the illness.

TABLE 6. Average rankings of the algorithms (Friedman).

Algorithm	Ranking
AHFSE	**1.25**
Random Forest	2.25
Adaboost	3.125
Bagging	3.75
Random Subspace	4.625

B. RESULTS

In our experiments, we applied mRMR, Fisher, Chi-Square and Relief-F as feature selection which will be utilized in our framework. The Levenberg-Marquardt back-propagation neural network was used as the fitness function learner and weak learner in the bagging process. The proposed framework approach was measured by the average accuracy of the dataset. Five-fold crossover validation was applied to reduce the randomness effects, and the experiments is repeated six times for each classification tasks. In here, Table 1 lists important parameters in our experiments.

In the following experiments, first, we explore the effect of adaptive hybrid feature selection within bagging process. Then, we compared our framework with several perspectives, such as: 1) comparison with other feature selection; 2) comparison with other ensemble methods; 3) comparison with other seizure detection using DWT as preprocessing step; finally, 4) comparison with other seizure detection using other preprocessing step. Detailed non-parametric test is also provided at supplementary materials in respective matter.

TABLE 7. Comparison with other seizure detection using DWT as preprocessing.

Authors	Method	Data selection	task type	acc	sen	spec
Guo et.al [18]	DWT with genetic programming	40% training - 60% testing	E-A E-A-D	99.2 93.5	- -	- -
Tawfik et.al [13]	DWT integrated with weighted - permutation entropy and SVM	10-fold cross validation	E-A E-B E-C E-D	91 82.5 84.5 95	- - - -	- - - -
Orhan et.al [29]	DWT with MLP and K-means - as probability distributions - of wavelet coefficient	unspecified fold - cross validation	E-A E-ABCD E-A-D E-AB-CD	**100** 99.6 **96.67** 95.6	100 100 94.12 92.38	100 98.4 97.98 97.93
Ubeyli [31]	DWT with mixture of - expert network	unspecified fold - cross validation	E-A-D	93.17	93	94
Acharya et.al [32]	DWT with fuzzy classifier	10-fold cross validation	E-A	99.7	100	100
Kumar et.al [30]	DWT with neural network based on - Approximate entropy	5-fold cross validation	E-A E-B E-C E-D E-ABCD	**100** 92.5 **100** 95 94	- - - - -	- - - - -
Our method	DWT and adaptive hybrid Feature - selection within Bagging with MLP	5-fold cross validation	E-A E-B E-C E-D E-ABCD E-A-D E-AB-CD	**100** **99.5** 98.5 **97.33(0.24)** **99(0.06)** 96(0.13) **97(0.28)**	100 100 98.83 97.76 97.32 98.27 96.53	100 99.01 98.1 97.49 99.26 98.87 98.93

1) EFFECT OF ADAPTIVE HYBRID FEATURE SELECTION WITHIN BAGGING

Table 2 contains optimum parameters for AHFSE. In here, due to we using four feature selection (mRMR, Fisher, Chi-Square and Relief-F), then length of individual is set to four. We also set the population number as twice of feature selection algorithm minus one. The reason is we don't need many initial population due to the individual will also obtained in population generation (iteration). Thus, by applied rank-aggregation within genetic algorithm, every bootstrap are guaranteed to obtain optimum feature subsets to generate new compact sample subsets. From Table 3 in part of AHFSE compared with None feature selection, can be seen that AHFSE with those parameters could obtain feature subsets within bagging process which improved the accuracy, sensitivity and specificity.

2) COMPARISON WITH OTHER FEATURE SELECTION

In here, we compared our framework with other feature selection. the following experiment are: our framework utilized adaptive hybrid feature selection within bagging process, whereas the other feature selection that are chosen (mRMR, Fisher, Chi-Square, Relief-F) is utilized in EEG training data and put into simple Bagging. From Table 3, in point of view 10 subsets features, our framework successfully obtained high performance in 7 of 8 classification tasks. Our framework is not better than Chi-Square in 'E-A-D' task type (where AHFSE obtained 93.67 and Chi-Square obtained 95). In point of view all subsets features, our framework obtained high performance in 5 of 8 classification tasks. From Table 4, our framework need 10 subset features to obtained desirable performance and confirm the presence of statistical difference with other feature selection regarding the Bagging method. The possible reason our framework obtain higher performance rather than simple feature selection due to AHFSE within rank-aggregation found optimum feature subsets which cover a larger set of possible underlying data

3) COMPARISON WITH OTHER ENSEMBLE METHODS

In here, other ensemble methods is compared with our optimum framework. As worth to mention, other ensemble methods did not applied any feature selection, thus all features in other ensemble methods is set to 45. From Table 5, our framework obtained high performance in 7 of 8 classification task. Adaboost successfully obtained high accuracy in 'E-C' and 'E-D' task type, but failed at multi-class problem such as 'E-A-D' and 'E-AB-CD'. Even though Adaboost successfully classified the seizure and seizure free, it not suitable for detection where the patient is not having seizure. Overall, our framework obtained higher ranking than other ensemble methods, which means it confirms the presence of significant statistical difference with other ensemble frameworks.

4) COMPARISON WITH OTHER SEIZURE DETECTION USING DWT AS PREPROCESSING

In here, we compared our framework with other seizure detection using DWT as preprocessing. From Table 7, our framework obtain high performance in 5 of 7 task-type. In 'E-A-D' task, our framework is comparable with Orhan *et al.* [29]. In 'E-C' task type, Kumar *et al.* [30] obtain satisfactory result than our framework. Thus, we can obtain conclusion that our framework obtained satisfactory results when applied together with DWT as preprocessing.

TABLE 8. Comparison with other seizure detection using other preprocessing.

Authors	Method	Data selection	task type	acc	sen	spec
Li et.al [9]	DWT based envelope analysis - with neural network	50% training - 50% testing	E-A-D	**98.78**	-	-
Supriya et.al [34]	Weighted visibility graph and SVM	training & testing	E-A	**100**	-	-
			E-B	97.25	-	-
			E-C	98.25	-	-
			E-D	93.25	-	-
Li et.al [14]	DT-CWT with SVM	10-fold cross validation	E-A-D	**98.87**	98.2	100
Zhang & Chen [33]	LMD based and SVM with - optimized by GA	50% training - 50% testing	E-A	**100**	100	100
			E-D	98.1(1.197)	-	97.4
			E-ABCD	98.87(0.63)	-	97.3
			E-A-D	98.47(0.63)	98.4	-
			E-AB-CD	98.4(0.56)	98.67	-
Sharma & Pachori [5]	EMD based on phase space - representation with Least Square SVM	10-fold cross validation	E-CD	98.67	100	96
Kaya et.al [35]	One dimensional of local binary - pattern bayes-net	10-fold cross validation	E-A	97.5	96	99
			E-D	99.5	99	100
			E-CD	95.5	95	96
			E-A-D	95.67	-	-
Tiwari et.al [36]	Key-point based local binary - pattern with SVM	10-fold cross validation	E-CD	99.45(0.25)	-	-
			E-ABCD	99.31(0.17)	-	-
			E-AB-CD	98.8(0.11)	-	-
Hassan et.al [23]	Tunable-Q wavelet transform - with decision tree bagging	out of bag	E-A	**100**	100	100
			E-C	**100**	100	100
			E-D	**100**	100	100
			E-ABCD	**99.6**	99.49	100
			E-A-D	98.67	96.67	98
			E-AB-CD	98.4	98.3	98.6
Sharma & Pachori [37]	Tunable-Q wavelet transform - with Least Square-SVM and - fractal dimension	10-fold cross validation	E-A	**100**	100	100
			E-B	**100**	100	100
			E-C	**100**	100	100
			E-D	99.5	100	99
			E-CD	**99.67**	100	99
			E-ABCD	**99.6**	100	98
Samiee et.al [6]	Rational discrete - short-time fourier transform and MLP	50% training - 50% testing	E-A	**100**	100	100
			E-B	99.3	99.6	99
			E-C	98.5	99.3	97.7
			E-D	94.9	95.6	94.1
			E-ABCD	98.1	99.2	93.8
Peker et.al [10]	Dual tree complex wavelet - transform with neural network - using complex value	10-fold cross validation	E-A	**100**	100	100
			E-ABCD	99.15	100	97.89
			E-A-D	**99.3**	99.4	98.8
			E-AB-CD	98.28	98.91	98.28
Our method	DWT and adaptive hybrid Feature - selection within Bagging with MLP	5-fold cross validation	E-A	**100**	100	100
			E-B	99.5	100	99.01
			E-C	98.5	98.83	98.1
			E-D	97.33(0.24)	97.76	97.49
			E-CD	98.44(0.16)	97.46	98.99
			E-ABCD	99(0.06)	97.32	99.26
			E-A-D	96(0.13)	98.27	98.87
			E-AB-CD	97(0.28)	96.53	98.93

5) COMPARISON WITH OTHER SEIZURE DETECTION USING OTHER PREPROCESSING

In here, we compared our framework with other seizure detection using other preprocessing methods. The reason we applied DWT as preprocessing step due to it is faster and widely used in epileptic seizure research. Because our research is focused in identify a framework and algorithm with better performance, we did not consider to obtain the perfect signal which could improve the detection results. Thus, several researchers applied different preprocessing which alleviate the problem of DWT such as: (1) Tunable-Q factor wavelet transform (TQWT) which modified version of DWT is applied by [5] and [23]. TQWT have several advantages such as: *a*) rational transfer functions that easily appointed properly in frequency domain, *b*) TQWT able to perfectly reconstruct the wavelet transform property, *c*) TQWT is able to tuning Q-factor, Q-factor is the value of oscillatory behavior, which is if the signal do not have oscillatory behavior then the wavelet transform which will used to analyze those signal supposed to be have low Q-factor values. (2) DWT with Envelope analysis using hilbert transform [9] which able to generate clearer and smoother curves also gives satisfactory results. (3) Local Mean Decomposition (LMD) [33] which transforms EEG signals into multiple product functions and applied mirror extending approach to tackle the effect of edge resulted in more reliable than traditional

Empirical Mode Decomposition(EMD). (4) Dual-tree Complex in Wavelet Transformation (DTCWT) is reported have many advantages than DWT [10], such as: *a*) DTCWT successful in selecting diagonal characteristic and used complex function instead relying on real-valued functions of main wavelet, *b*) DTCWT more successful than DWT in differentiating the input-signal changes.

From description above, simple preprocessing is not enough to obtain satisfactory results. Thus, in next research, the modification of traditional preprocessing is required to applied as preprocessing step.

V. CONCLUSION & FUTURES WORKS

We propose the AHFSE for epileptic seizure classification. Compared to traditional bagging framework approaches, our AHFSE framework is characterized as follows: 1) the adaptive hybrid (multiple) feature selection algorithm selects the most representative features by means of rank aggregation to avoid redundancy and improve the performance of the classifier; 2) the AHFSE algorithm is used within the bagging process to obtain new compact sample subsets to feed into the classifier; and 3) the voting method is used to obtain the final detection and classification results.

Two main highlights of the experiments are as follows. First,: 1) using AHFSE to obtain new samples within the bagging process improves the specificity, sensitivity or accuracy of the selected features compared to traditional methods. 2) the adaptive process helps the framework obtain the optimum feature selection combination.

Conclusion that can be drawn from this research is: First, the AHFSE algorithm combines several feature selections and obtains the single-ranking features playing important roles in the performance of seizure detection and classification. Second, our framework obtain satisfactory results when compared with: (1) other feature selection; (2) other ensemble methods; (3) other seizure detection research using discrete wavelet transform (DWT) as preprocessing methods. Third, modification of traditional preprocessing is required to applied as preprocessing step to obtain more satisfactory results.

As explained in previous section, our algorithm needs longer time to obtain compact sample subsets, thus in future works, particle swarm optimization can be used as part of AHFSE replacing the genetic algorithm due to simplicity (which is not using combination and mutation) resulted in computational cost-effectiveness. Classifier ensemble pruning (reduction) also can be adopted to reduce the runtime overheads [38], [39]. Also, in preprocessing directions, necessity to applied modification of traditional preprocessing is required. Thus, several modifications of traditional preprocessing will be considered.

REFERENCES

[1] R. S. Fisher *et al.*, "Epileptic seizures and epilepsy: Definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, 2005, doi: 10.1111/j.0013-9580.2005.66104.x.

[2] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," Ph.D. dissertation, Division Health Sci. Technol., Massachusetts Inst. Technol., Cambridge, MA, USA, 2009. [Online]. Available: http://hdl.handle.net/1721.1/54669

[3] A. Bhattacharyya, R. B. Pachori, A. Upadhyay, and U. R. Acharya, "Tunable-Q wavelet transform based multiscale entropy measure for automated classification of epileptic EEG signals," *Appl. Sci.*, vol. 7, no. 4, p. 385, 2017, doi: 10.3390/app7040385.

[4] D. Bhati, M. Sharma, R. B. Pachori, and V. M. Gadre, "Time–frequency localized three-band biorthogonal wavelet filter bank using semidefinite relaxation and nonlinear least squares with epileptic seizure EEG signal classification," *Digit. Signal Process.*, vol. 62, pp. 259–273, Mar. 2017, doi: 10.1016/j.dsp.2016.12.004.

[5] R. Sharma and R. B. Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1106–1117, 2015, doi: 10.1016/j.eswa.2014.08.030.

[6] K. Samiee, P. Kovacs, and M. Gabbouj, "Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 541–552, Feb. 2015, doi: 10.1109/tbme.2014.2360101.

[7] D. Bhati, R. B. Pachori, and V. M. Gadre, "A novel approach for time–frequency localization of scaling functions and design of three-band biorthogonal linear phase wavelet filter banks," *Digit. Signal Process.*, vol. 69, pp. 309–322, Oct. 2017, doi: 10.1016/j.dsp.2017.07.008.

[8] E. Abdulhay, V. Elamaran, M. Chandrasekar, V. S. Balaji, and K. Narasimhan, "Automated diagnosis of epilepsy from EEG signals using ensemble learning approach," *Pattern Recognit. Lett.*, to be published, doi: 10.1016/j.patrec.2017.05.021.

[9] M. Li, W. Chen, and T. Zhang, "Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble," *Biomed. Signal Process. Control*, vol. 31, pp. 357–365, Jan. 2017, doi: 10.1016/j.bspc.2016.09.008.

[10] M. Peker, B. Sen, and D. Delen, "A novel method for automated diagnosis of epilepsy using complex-valued classifiers," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 1, pp. 108–118, Jan. 2016, doi: 10.1109/jbhi.2014.2387795.

[11] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowl.-Based Syst.*, vol. 45, pp. 147–165, Jun. 2013, doi: 10.1016/j.knosys.2013.02.014.

[12] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K.-H. Ng, and J. S. Suri, "Automated diagnosis of epileptic EEG using entropies," *Biomed. Signal Process. Control*, vol. 7, no. 4, pp. 401–408, Jul. 2012, doi: 10.1016/j.bspc.2011.07.007.

[13] N. S. Tawfik, S. M. Youssef, and M. Kholief, "A hybrid automated detection of epileptic seizures in EEG records," *Comput. Elect. Eng.*, vol. 53, pp. 177–190, Jul. 2016, doi: 10.1016/j.compeleceng.2015.09.001.

[14] M. Li, W. Chen, and T. Zhang, "Automatic epileptic EEG detection using DT-CWT-based non-linear features," *Biomed. Signal Process. Control*, vol. 34, pp. 114–125, Apr. 2017, doi: 10.1016/j.bspc.2017.01.010.

[15] F. Alzami, D. Wang, Z. Yu, J. You, H.-S. Wong, and G. Han, "Robust epileptic seizure classification," in *Intelligent Computing Theories and Application*. Cham, Switzerland: Springer, 2016, pp. 363–373, doi: 10.1007/978-3-319-42294-7_32.

[16] S.-H. Lee, J. S. Lim, J.-K. Kim, J. Yang, and Y. Lee, "Classification of normal and epileptic seizure EEG signals using wavelet transform, phase-space reconstruction, and Euclidean distance," *Comput. Methods Programs Biomed.*, vol. 116, no. 1, pp. 10–25, Aug. 2014, doi: 10.1016/j.cmpb.2014.04.012.

[17] E. Pippa *et al.*, "Improving classification of epileptic and non-epileptic EEG events by feature selection," *Neurocomputing*, vol. 171, pp. 576–585, Jan. 2016, doi: 10.1016/j.neucom.2015.06.071.

[18] L. Guo, D. Rivero, J. Dorado, C. R. Munteanu, and A. Pazos, "Automatic feature extraction using genetic programming: An application to epileptic EEG classification," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10425–10436, Aug. 2011, doi: 10.1016/j.eswa.2011.02.118.

[19] I. Guler and E. D. Ubeyli, "Multiclass support vector machines for EEG-signals classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 2, pp. 117–126, Mar. 2007.

[20] S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr, "Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 512–518, Feb. 2008, doi: 10.1109/tbme.2007.905490.

[21] Z. Yu, H. Chen, J. You, G. Han, and L. Li, "Hybrid fuzzy cluster ensemble framework for tumor clustering from biomolecular data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 10, no. 3, pp. 657–670, May 2013, doi: 10.1109/tcbb.2013.59.

[22] Z. Yu, L. Li, J. Liu, and G. Han, "Hybrid adaptive classifier ensemble," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 177–190, Feb. 2015, doi: 10.1109/tcyb.2014.2322195.

[23] A. R. Hassan, S. Siuly, and Y. Zhang, "Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating," *Comput. Methods Programs Biomed.*, vol. 137, pp. 247–259, Dec. 2016, doi: 10.1016/j.cmpb.2016.09.008.

[24] R. Kolde, S. Laur, P. Adler, and J. Vilo, "Robust rank aggregation for gene list integration and meta-analysis," *Bioinformatics*, vol. 28, no. 4, pp. 573–580, Jan. 2012, doi: 10.1093/bioinformatics/btr709.

[25] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[26] H. Liu and H. Motoda, Eds., *Computational Methods of Feature Selection*. London, U.K.: Chapman & Hall, 2008.

[27] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes," in *Proc. 7th IEEE Int. Conf. Tools Artif. Intell.*, J. Vassilopoulos, Ed. Herndon, Virginia: IEEE Computer Society, Nov. 1995, pp. 388–391.

[28] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 6, p. 061907, Nov. 2001, doi: 10.1103/physreve.64.061907.

[29] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13475–13481, Sep. 2011, doi: 10.1016/j.eswa.2011.04.149.

[30] Y. Kumar, M. L. Dewal, and R. S. Anand, "Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network," *Signal, Image Video Process.*, vol. 8, no. 7, pp. 1323–1334, Aug. 2012, doi: 10.1007/s11760-012-0362-9.

[31] E. D. Übeyli, "Wavelet/mixture of experts network structure for EEG signals classification," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1954–1962, Apr. 2008, doi: 10.1016/j.eswa.2007.02.006.

[32] U. R. Acharya, S. V. Sree, P. C. A. Ang, R. Yanti, and J. S. Suri, "Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals," *Int. J. Neural Syst.*, vol. 22, no. 2, p. 1250002, Apr. 2012, doi: 10.1142/s0129065712500025.

[33] T. Zhang and W. Chen, "LMD based features for the automatic seizure detection of EEG signals using SVM," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1100–1108, Aug. 2017, doi: 10.1109/tnsre.2016.2611601.

[34] S. Supriya, S. Siuly, H. Wang, J. Cao, and Y. Zhang, "Weighted visibility graph with complex network features in the detection of epilepsy," *IEEE Access*, vol. 4, pp. 6554–6566, 2016, doi: 10.1109/access.2016.2612242.

[35] Y. Kaya, M. Uyar, R. Tekin, and S. Yıldırım, "1D-local binary pattern based feature extraction for classification of epileptic EEG signals," *Appl. Math. Comput.*, vol. 243, pp. 209–219, Sep. 2014, doi: 10.1016/j.amc.2014.05.128.

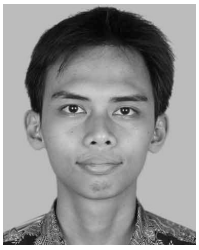
[36] A. K. Tiwari, R. B. Pachori, V. Kanhangad, and B. K. Panigrahi, "Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 4, pp. 888–896, Jul. 2017, doi: 10.1109/jbhi.2016.2589971.

[37] M. Sharma and R. B. Pachori, "A novel approach to detect epileptic seizures using a combination of tunable-Q wavelet transform and fractal dimension," *J. Mech. Med. Biol.*, vol. 17, no. 7, p. 1740003, Nov. 2017, doi: 10.1142/s0219519417400036.

[38] R. Diao, F. Chao, T. Peng, N. Snooke, and Q. Shen, "Feature selection inspired classifier ensemble reduction," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1259–1268, Aug. 2014, doi: 10.1109/tcyb.2013.2281820.

[39] G. Yao, H. Zeng, F. Chao, C. Su, C.-M. Lin, and C. Zhou, "Integration of classifier diversity measures for feature selection-based classifier ensemble reduction," *Soft Comput.*, vol. 20, no. 8, pp. 2995–3005, Nov. 2015, doi: 10.1007/s00500-015-1927-7.

[40] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2001.



FARRIKH ALZAMI is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include machine learning, classification, clustering, and ensemble learning.

JUAN TANG received the master's degree from Sun Yat-sen University. She is currently a Neonatologist and a Technologist-in-Charge of clinical medicine inspection technology with the Guangzhou Women and Children's Medical Centre. The main research area is neonatal brain injury, EEG, electroencephalogram, and ultrasound imaging.

ZHIWEN YU (S'06–M'08–SM'14) received the Ph.D. degree from the City University of Hong Kong in 2008. He is currently a Professor with the School of Computer Science and Engineering, South China University of Technology, and an Adjunct Professor with Sun Yat-sen University. He has published over 100 referred journal papers and international conference papers, including the TKDE, TEVC, TCYB, TMM, and TCVST. His research areas of interest include data mining, machine learning, and pattern recognition. He is a Distinguished Council Member of the China Computer Federation and a Senior Member of the ACM and CAAI.

SI WU received the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2013. He is currently an Associate Professor with the School of Computer Science and Engineering, South China University of Technology, China. His research interests include computer vision and pattern recognition.

C. L. PHILIP CHEN (S'88–M'88–SM'94–F'07) received the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1985, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, 1988, . He is currently a Chair Professor with the Department of Computer and Information Science and the Dean of the Faculty of Science and Technology, University of Macau, Macau, China. His current research interests include computational intelligence, systems, and cybernetics.

JANE YOU is currently a Professor with the Department of Computing, The Hong Kong Polytechnic University (PolyU), and the Chair of the Department Research Committee. She has worked extensively in the fields of image processing, medical imaging, computer-aided diagnosis, and pattern recognition. So far, she has published over 190 research papers with over 1,000 nonself-citations. She has served as the principal investigator for one ITF project, three GRF projects, and many other joint grants since she joined PolyU in 1998. She is also a team member for two successful patents (one Hong Kong patent and one U.S. patent) and three awards, including the Hong Kong Government Industrial Awards. Her work on retinal imaging earned her a Special Prize and Gold Medal with Jury's Commendation at the 39th International Exhibition of Inventions of Geneva in 2011 and second place at an international competition (SPIE Medical Imaging'2009 Retinopathy Online Challenge in 2009). Her research output on retinal imaging has successfully led to technology transfer with clinical applications. She is also an Associate Editor for *Pattern Recognition* and other journals.

JUN ZHANG (M'02–SM'08–F'16) received the Ph.D. degree in electrical engineering from the City University of Hong Kong, Hong Kong, in 2002. Since 2016, he has been with the South China University of Technology, Guangzhou, China, where he is currently a Cheung Kong Professor. He has authored seven research books and book chapters, and over 100 technical papers in his research areas. His current research interests include computational intelligence, cloud computing, big data, and wireless sensor networks. He was a recipient of the First-Grade Award in Natural Science Research from the Ministry of Education, China, in 2009, and the China National Funds for Distinguished Young Scientists from the National Natural Science Foundation of China in 2011. He is currently an Associate Editor for the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, and the IEEE TRANSACTIONS ON CYBERNETICS. He is the Founding and Current Chair of the IEEE Guangzhou Subsection and the Founding and Current Chair of ACM Guangzhou Chapter.