7th International Conference on Intelligent Control and Information Processing
December 1-4, 2016; Siem Reap, Cambodia

# Efficient Feature Extraction Framework for EEG Signals Classification

Weijie Ren, and Min Han
Faculty of Electronic Information and Electrical Engineering
Dalian University of Technology
Dalian, China
Email: minhan@dlut.edu.cn

Jun Wang
Department of Computer Science
City University of Hong Kong
Hong Kong, China
Email: jwang.cs@cityu.edu.hk

Dan Wang
Marine Engineering College
Dalian Maritime University
Dalian, China
Email: dwang@dlmu.edu.cn

Tieshan Li
Navigation College
Dalian Maritime University
Dalian, China
Email: tieshanli@126.com

***Abstract*—Feature extraction and classification for EEG signals are key technologies in medical applications. This paper proposes an efficient feature extraction framework that combines hybrid feature extraction and feature selection method. In order to fully exploit information from EEG signals, several feature extraction methods of different types are applied, which are autoregressive model, discrete wavelet transform, wavelet packet transform and sample entropy. After information fusion, feature selection methods are introduced to deal with redundant and irrelevant information, which is advantageous to classification. In this phase, global optimization strategy based on binary particle swarm optimization (BPSO) is presented to enhance the performance of feature selection. To evaluate the results of feature extraction, experiments of class separability are conducted. Classification results on EEG dataset of university of Bonn show the superiority of the proposed method.**

***Keywords—EEG signals classification; feature extraction; feature selection; feature transformation***

## I. INTRODUCTION

Electroencephalogram (EEG) records electrical activity of the brain through electrodes on the scalp. EEG can reflect the physiological functions of the brain, which is usually used for analyzing and diagnosing diseases [1]. In the past, due to the limitation of technology, people analyzed EEG signals through visual analysis, which is usually unreliable and time-consuming. In recent years, with the development of computer science and artificial intelligence technique, automatic classification of EEG signals is becoming research focus around the world [2].

With regard to automatic classification of EEG signals, EEG feature extraction is the question of fundamental importance [3]. EEG is a kind of random signals that contains highly complex information. Thus, single feature extraction approach usually cannot describe the property of EEG signals completely [4]. To solve this problem, many scholars have proposed combined methods for EEG feature extraction, which include different categories of characteristics [5, 6]. Hence, it is necessary to take full advantage of different types of feature extraction methods and explore the optimal combination of input features.

While a wealth of information is extracted from original EEG signals, the curse of dimensionality appears. Moreover, many of them might be redundant or irrelevant features, which have negative effects on classification task. Therefore, dimensionality reduction [7] is an important preprocessing step in EEG signals classification. The goal of dimensionality reduction is to remove the redundant and irrelevant information, simplify classification model and enhance generalization ability. Feature transformation and feature selection are two popular methods. Subasi et al. [8] proposed using PCA, ICA and LDA to reduce the dimensionality of data and improve the generalization performance. Besides, many researches have been done for feature selection. Statistical test [1], genetic algorithm [9], Fisher criterion [10], linear regression [11] etc. are widely used for processing EEG data and have achieved good results.

In this paper, we present an efficient feature extraction framework for EEG signals classification. Hybrid feature extraction can well express complex characteristics of EEG signals. Furthermore, feature selection based on global optimization strategy improves the class separability of EEG samples. The paper is organized as follows: section 2 provides the hybrid feature extraction method, section 3 gives a detail presentation for feature transformation and feature selection algorithms, global search strategy for feature selection is presented in section 4, the experimental results on EEG dataset of university of Bonn are analyzed in section 5, and section 6 gives the conclusions.

## II. HYBRID FEATURE EXTRACTION

Feature extraction is an essential condition for automatic classification of EEG signals. At present, commonly used feature extraction methods mainly include time domain analysis, frequency domain analysis, time-frequency domain analysis and nonlinear analysis methods. In this paper, we apply different kinds of methods simultaneously, which can well express complex characteristics of EEG signals. The typical feature extraction algorithms used in our research is shown as follows:

- Autoregressive (AR) model: AR model is a powerful time domain analysis method for signal processing. Since AR model fits the time series based on its own history information, feature information of EEG signals can be derived from AR coefficients. In practical applications, the order of AR model is a key parameter,

This work was supported by the project (61374154) of the National Natural Science Foundation of China.

which is determined by Akaike information criterion in this paper.

- Discrete wavelet transform (DWT): DWT is a popular time-frequency domain analysis method for signal processing. The signal is decomposed using high-pass and low-pass filters simultaneously. It provides an efficient hierarchical framework for multiresolution analysis of EEG signals. Wavelet coefficients can represent the time-frequency characteristics of EEG signals efficiently.
- Wavelet packet transform (WPT): WPT is developed from the theory of wavelet transform. WPT makes decomposition for low frequency and high frequency part at the same time, which creates a full binary tree. Compared to DWT, WPT provides more detail information. Statistical characteristics of wavelet packet coefficients are used for EEG signals classification.
- Sample entropy (SampEn): Sample entropy is a nonlinear dynamics parameter that describes the complexity of physiological time series. The more complex the time series is, the greater the value of sample entropy is. As sample entropy is easy to implement and has strong anti-noise capacity, it is suitable for assessing real EEG signals.

## III. Feature Transformation & Feature Selection

Dimensionality reduction is one of the basic problems in pattern recognition and classification. The main purpose of feature transformation and feature selection is to remove redundant or irrelevant information and improve classification performance. In this paper, we prefer to use principal component analysis for feature transformation, and design efficient feature selection algorithms based on mutual information and Fisher score criterion.

### A. Principal Component Analysis

Principal component analysis (PCA) [8] is a statistical learning method, which converts a set of correlated features into linearly uncorrelated features based on orthogonal transformation. The features obtained by PCA are called principal components, of which the number is less than or equal to the number of original features. Therefore, PCA can remove unnecessary features and retain original information as much as possible. Besides, PCA transformation could project high-dimensional data onto low-dimensional subspace, which displays the most informative viewpoint for dataset.

Mathematically, we define the original data $\boldsymbol{X} \in \mathbb{R}^{d\times n}$ with $d$-dimensional vectors. The PCA transformation is conducted by $\boldsymbol{w}_i = [w_{i1}, w_{i2}, \ldots, w_{id}] \in \mathbb{R}^d, i = 1, \ldots, d$ , by which the new principal component vectors are obtained

$$\boldsymbol{T}_L = \boldsymbol{W}_L \boldsymbol{X} \tag{1}$$

where $L \le d$ is the number of principal components, $\boldsymbol{W}_L = [\boldsymbol{w}_1; \ldots; \boldsymbol{w}_L] \in \mathbb{R}^{L\times d}$ and $\boldsymbol{T}_L \in \mathbb{R}^{L\times n}$.

### B. Minimum Redundancy Maximum Relevance

Minimal redundancy maximal relevance (mRMR) [12] criterion is a widely used feature selection algorithm based on mutual information, which considers the relevance and redundancy simultaneously. Given the training set $\{\boldsymbol{X}, \boldsymbol{y}\}$ with $n$ samples and $d$-dimensional features, the input data $\boldsymbol{X} \in \mathbb{R}^{d\times n}$ is reduced to $\boldsymbol{Z} \in \mathbb{R}^{m\times n}$ after feature selection. The objective function of mRMR is shown as follows:

$$\max \left\{ \frac{1}{m} \sum_{X_i \in Z} I(\boldsymbol{X}_i; \boldsymbol{y}) - \frac{1}{m^2} \sum_{X_i, X_j \in Z} I(\boldsymbol{X}_i; \boldsymbol{X}_j) \right\} \tag{2}$$

where $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are the input features, $\boldsymbol{y}$ is the target output and $m$ is the number of selected features. $I(\boldsymbol{X}_i; \boldsymbol{y})$ is the mutual information of $\boldsymbol{X}_i$ and $\boldsymbol{y}$:

$$I(\boldsymbol{X}_i; \boldsymbol{y}) = \sum p(\boldsymbol{X}_i, \boldsymbol{y}) \log \frac{p(\boldsymbol{X}_i, \boldsymbol{y})}{p(\boldsymbol{X}_i) p(\boldsymbol{y})} \tag{3}$$

In order to increase the searching accuracy, the criterion can be formulated as global optimization problem [13]. We introduce a feature vector $\boldsymbol{w} \in \mathbb{R}^d$ , in which the element represents whether the feature is selected or not. Then, the criterion is converted to

$$\begin{aligned} &\max \ \boldsymbol{w}^{\mathrm{T}} \boldsymbol{c} - \frac{1}{2m} \boldsymbol{w}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{w} \\ &\text{s.t.} \ \boldsymbol{w}_i \in \{0, 1\}, i = 1, \ldots, d, \|\boldsymbol{w}\|_1 = m \end{aligned} \tag{4}$$

where $\boldsymbol{c}_i = I(\boldsymbol{X}_i; \boldsymbol{y})$, $\boldsymbol{Q}_{ij} = I(\boldsymbol{X}_i; \boldsymbol{X}_j), i \ne j$ , and $\boldsymbol{X}_i, \boldsymbol{X}_j \in \boldsymbol{X}$ .

### C. Fisher Score

Fisher score [10] is a kind of commonly used supervised feature selection algorithm. The basic idea is to find a feature subset, which satisfies maximizing between-class distances and minimizing within-class distances of samples. The expression of Fisher score is shown below

$$F(\boldsymbol{Z}) = tr\left\{ \boldsymbol{S}_b \boldsymbol{S}_w^{-1} \right\} \tag{5}$$

where $tr\{\cdot\}$ denotes the trace of matrix. The between-class scatter matrix $\boldsymbol{S}_b$ and within-class scatter matrix $\boldsymbol{S}_w$ are defined as

$$\boldsymbol{S}_b = \sum_{i=1}^{c} n_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^{\mathrm{T}} \tag{6}$$

$$\boldsymbol{S}_w = \sum_{i=1}^{c}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij} - \boldsymbol{\mu}_i)(\boldsymbol{z}_{ij} - \boldsymbol{\mu}_i)^{\mathrm{T}} \tag{7}$$

where $\boldsymbol{\mu}$ is average value of all samples, $\boldsymbol{\mu}_i$ is average value of samples in the $i$th class, $n_i$ is the number of samples in the $i$th class, and $c$ is the number of classes. In order to facilitate solving the algorithm, feature selection based on Fisher score can also be formulated as a global optimization problem:

$$\begin{aligned} &\max \ \operatorname{tr}\left\{\left(\operatorname{diag}(\boldsymbol{w})\boldsymbol{S}_b\operatorname{diag}(\boldsymbol{w})\right)\left(\operatorname{diag}(\boldsymbol{w})\boldsymbol{S}_w\operatorname{diag}(\boldsymbol{w})\right)^{-1}\right\} \\ &\text{s.t. } \ w_i \in \{0,1\},\, i = 1,\ldots,d,\, \|\boldsymbol{w}\|_1 = m \end{aligned} \tag{8}$$

where $\operatorname{diag}(\boldsymbol{w})$ is a diagonal matrix and $\boldsymbol{w}$ is its primary diagonal element.

## IV. Global Searching Strategy

Search strategy is a key point in feature selection problem. In order to obtain global optimal solution, the criteria of feature selection have been transformed into global optimization problems, which are shown in (4) and (8). Considering the computational accuracy and efficiency, a popular evolutionary algorithm is exploited in this research, that is binary particle swarm optimization (BPSO) algorithm.

Based on the basic PSO algorithm, Kennedy et al [14] proposed binary PSO to solve the 0-1 integer programming problem. For each particle with $d$ elements, the velocity expresses as $\boldsymbol{V} = (V_1, V_2, \ldots, V_d)$ and the position expresses as $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$. In BPSO, the each dimension of particle's position is limited to 0 or 1. The update formulas of velocity and position are shown as follows

$$V_i = w \times V_i + c_1 r_1 \left(\text{pbest}_i - X_i\right) + c_2 r_2 \left(\text{gbest}_i - X_i\right) \tag{9}$$

$$X_i = \begin{cases} 1 & \text{rand}() < \operatorname{sig}(V_i) \\ 0 & \text{else} \end{cases} \tag{10}$$

where $w$ is inertia weight, $c_1$ and $c_2$ are accelerated factors, $r_1$ and $r_2$ are random numbers in $[0,1]$, $\text{pbest}_i$ denotes the optimal value of individual and $\text{gbest}_i$ denotes the global optimal value. $\operatorname{sig}(\cdot)$ is S-shaped transfer function and rand() is a random number between 0 and 1. However, while the velocity $V_i$ is close to zero, the position $X_i$ might still change, which may lead to non-convergent. Therefore, we consider V-shaped transfer function instead of S-shaped [15], which is shown below

$$S'(V_i) = 2 \times \left|\operatorname{sig}(V_i) - 0.5\right| \tag{11}$$

Thus, the update formula of position is revised to

$$X_i = \begin{cases} \text{exchange}(X_i) & \text{rand}() < S'(V_i) \\ X_i & \text{else} \end{cases} \tag{12}$$

where $\text{exchange}(\cdot)$ expresses the change of particle's position.

Finally, the procedure of global feature selection based on BPSO is shown as follows:

**Step 1**: Initialization. Set basic parameters of BPSO algorithm, including population size, maximum number of iterations, inertia weight, accelerated factors and so on. Determine the size of feature subset.

**Step 2**: Encoding. Every particle contains $d$ elements, which represent $d$-dimension input features. Each element is randomly encoded as "0" or "1", where "1" means to select the corresponding feature.

**Step 3**: Constructing fitness function. Based on (4) and (8), we can obtain the object function to be optimized.

**Step 4**: Updating of velocity and position. According to value of fitness function, local and global optimal position can be derived. Then, update the velocity and position of all particles based on (9) and (12).

**Step 5**: Repeat step 4 until satisfy stop condition or reach maximum iterations. The global optimal solution is the feature selection result.

## V. Experiments and Results

We evaluate the performance of the proposed feature extraction framework based on the EEG dataset of university of Bonn. The dataset contains five subsets (A-E), which comes from healthy volunteers and epilepsy patients, respectively. In this experiment, we consider the multi-classification task for subsets A, D and E, which corresponds to healthy EEG signals, EEG signals from seizure free intervals and seizure activity, respectively. The procedure of experiment is shown in Fig. 1. Our method is compared with single feature extraction method and other related researches.

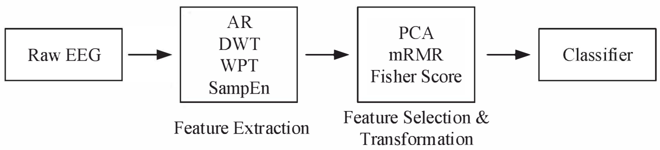


Fig. 1. The procedure of the proposed method for EEG signals classification.

### A. Evaluation of Feature Extraction

At the beginning, the input features of EEG samples are obtained by AR model, discrete wavelet transform, wavelet packet transform and sample entropy. The hybrid input feature vector is composed of 83 features. Then, we apply feature selection methods (mRMR and Fisher score) based on global search strategy to select useful feature information. A group of candidate subsets of different sizes can be obtained. In order to determine the optimal feature subset, we use cross validation procedure in this stage. The training set is divided into two parts, which are training subset and validation subset. The validation results are shown in Fig. 2. Finally, the subsets with maximum validation accuracy are selected for classification. Compared to the original feature set, the dimensionality of selected subset is obviously reduced.

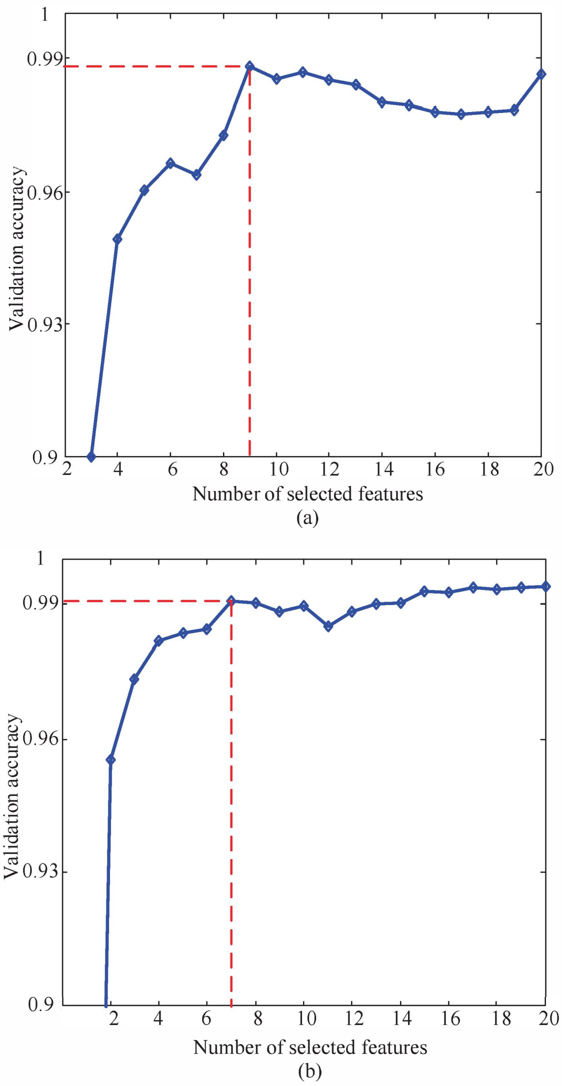


Fig. 2. Validation results for feature selection methods. (a) and (b) correspond to the results of mRMR and Fisher score feature selection.

To assess the results of feature extraction in visual sense, we project the EEG samples onto low-dimensional subspaces based on PCA transformation. Fig. 3 (a)-(c) shows the EEG samples of different classes in the most informative viewpoint, where subset A, D and E express in blue, red and yellow color, respectively. It is apparent that the samples obtained by hybrid feature extraction method are not separated completely and overlapped between different classes. Through feature selection, separability of EEG samples increases to a certain degree. Compared to others, EEG samples obtained by hybrid feature extraction with Fisher score feature selection can be distinguished clearly.

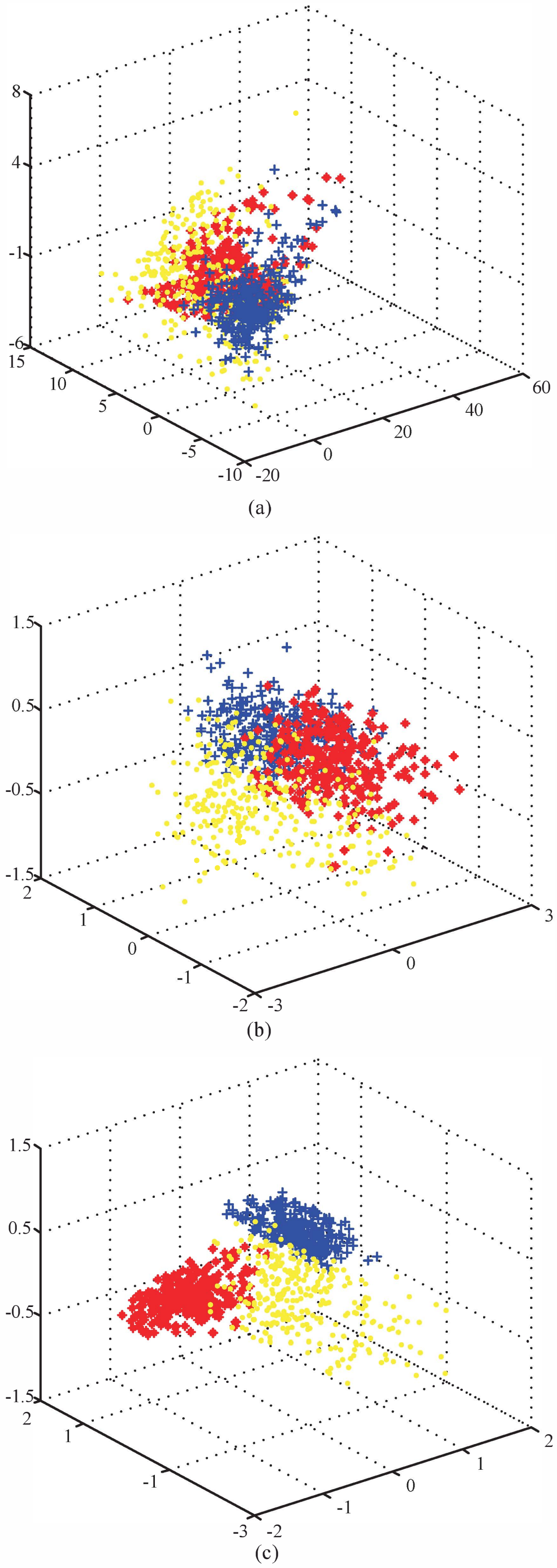


Fig. 3. Samples in low-dimensional subspaces based on PCA transformation. (a) corresponds to samples of hybrid features. (b) and (c) correspond to samples after mRMR and Fisher score feature selection.

TABLE I. COMPARISON OF CLASS SEPARABILITY BETWEEN DIFFERENT FEATURE EXTRACTION METHODS

| Feature extraction methods | $tr(\boldsymbol{S}_b)$ | $tr(\boldsymbol{S}_w)$ | $J$ |
|---|---|---|---|
| AR | 0.3727 | 0.4951 | 0.7528 |
| DWT | 15.0575 | 57.3295 | 0.2626 |
| WPT | 0.9963 | 1.5042 | 0.6624 |
| Hybrid features | 16.7494 | 59.3812 | 0.2821 |
| Hybrid features + PCA | 16.7473 | 59.0770 | 0.2835 |
| Hybrid features + mRMR | 0.8073 | 0.9615 | 0.8396 |
| Hybrid features + Fisher Score | 0.8556 | 0.5424 | 1.5776 |

Furthermore, quantitative analysis of feature extraction results is conducted. Class separability is an effective evaluation methodology for classification. It is calculated based on between-class and within-class scatter matrixes and the expression of class separability is defined in equation (5). In theory, the greater the class separability is, the better the ability of classification will be. Table 1 presents the class separability of EEG samples obtained by different methods. The hybrid features of several feature extraction methods achieves low class separability, because of interference of redundant and irrelevant features. Moreover, after PCA transformation, class separability have not been improved. By contrast, the class separability can be improved significantly by means of feature selection. Comparing with other approaches, the hybrid features with Fisher score feature selection achieve the highest class separability, which indicates the effectiveness of the proposed method.

## B. Classification Results

In this section, multi-classification experiments are conducted on EEG dataset of university of Bonn. After feature extraction, 1200 samples of three classes are obtained, of which 900 samples are randomly selected for training and the remaining are used for testing. Two classifiers are employed, namely decision tree (DT) and extreme learning machine (ELM). As the classification result of ELM is random, final classification result of ELM is the average value of 50 independent experiments. Table 2 shows the classification results of Bonn EEG dataset, including training and testing accuracies (%).

Experimental results show that the classification accuracies of single feature extraction methods are clearly lower than hybrid feature extraction method with the same classifier. AR model, DWT, WPT and SampEn cannot fully describe complex characteristics of EEG signals, which results in poor classification performance. On the contrary, hybrid features can express the EEG signals well, which benefits data classification. However, the high input dimensionality increases the computational burden.

The hybrid features contain 83 dimensionalities, which can be significantly decreased by feature selection and feature transformation. As class separability is not improved after PCA transformation, the classification result of hybrid feature extraction with PCA transformation is worse than feature

TABLE II. THE CLASSIFICATION RESULTS OF BONN EEG DATASET

| Feature extraction methods | No. of feature | DT | | ELM | |
|---|---|---|---|---|---|
| | | Train | Test | Train | Test |
| AR | 36 | 88.78 | 78.33 | 97.58 | 94.60 |
| DWT | 22 | 93.22 | 92.00 | 91.78 | 90.83 |
| WPT | 24 | 91.78 | 89.00 | 97.12 | 94.54 |
| SampEn | 1 | 75.00 | 67.00 | 73.97 | 71.46 |
| Hybrid features | 83 | 98.11 | 92.67 | 98.30 | 95.91 |
| Hybrid features + PCA | 9 | 92.33 | 86.00 | 99.24 | 96.80 |
| Hybrid features + mRMR | 9 | 95.89 | 94.33 | 99.14 | 98.31 |
| Hybrid features + Fisher Score | 7 | 97.00 | 94.67 | 99.69 | 98.95 |

TABLE III. COMPARISON OF RELATED RESEARCHES ON BONN EEG DATASET

| | Feature extraction | Classifier | Accuracy |
|---|---|---|---|
| Übeyli [16] | DWT | mixture of experts | 93.17 |
| Song [17] | Sample Entropy | ELM | 95.67 |
| Übeyli [18] | Lyapunov exponents | MLPNN | 96.33 |
| Güler [19] | Lyapunov exponents | KNN | 96.79 |
| Acharya [20] | DWT, ICA coefficients | SVM | 96.00 |
| Acharya [21] | Entropies | Fuzzy | 98.10 |
| Our method | AR, DWT, WPT and SampEn | ELM | 98.95 |

selection method. Through comparison and analysis of experimental results, the hybrid feature extraction with Fisher score feature selection achieves the highest classification accuracy of 98.95%. Under the condition of different classifiers, the proposed method gets better results than other feature extraction methods, which shows that the proposed method has good generalization ability in real world applications.

Table 3 shows the comparison of our method with other related researches. According to the same classification task, our method achieves higher classification accuracy than other researches. In comparison to feature extraction method of EEG signals, the proposed hybrid feature extraction framework has obvious advantages. In the further, we concentrate on expanding our method to other practical applications, which will play a role in auxiliary decision-making for analyzing human brain activities and diagnosing diseases.

# VI. CONCLUSIONS

We have presented a novel feature extraction framework for EEG signals classification. Aiming at feature extraction method can only express single property of EEG signals, different types of feature extraction methods are combined to fully exploit information from original EEG data. Feature selection method are investigated for selecting important features, which can significantly enhance classification capability. The experiments of class separability indicate that the proposed method is superior to other feature extraction methods. Classification results show that the proposed method performs well on EEG dataset of university of Bonn.

## REFERENCES


[1] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, and J. E. Koh, "Application of entropies for automated diagnosis of epilepsy using EEG signals: A review," Knowl.-based Syst., vol. 88, pp. 85-96, 2015.

[2] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," J. Neural Eng., vol. 4, no. 2, pp. R1-R13, 2007.

[3] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," IEEE Trans. Affect. Comput., vol. 5, no. 3, pp. 327-339, 2014.

[4] J. Atkinson, and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," Expert Syst. Appl., vol. 47, pp. 35-41, 2016.

[5] G. Rodríguez-Bermúdez, P. J. García-Laencina, J. Roca-González, and J. Roca-Dorda, "Efficient feature selection and linear discrimination of EEG signals," Neurocomputing, vol. 115, pp. 161-165, 2013.

[6] E. Gallego-Jutglà, J. Solé-Casals, F. B. Vialatte, M. Elgendi, A. Cichocki, and J. Dauwels, "A hybrid feature selection approach for the early diagnosis of Alzheimer's disease," J. Neural Eng., vol. 12, no. 1, pp. 016018, 2015.

[7] J. P. Cunningham, and Z. Ghahramani, "Linear dimensionality reduction: Survey, insights, and generalizations," J. Mach. Learn. Res., vol. 16, pp. 2859-2900, 2015.

[8] A. Subasi, and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," Expert Syst. Appl., vol. 37, no. 12, pp. 8659-8666, 2010.

[9] V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, "A new approach for EEG feature extraction in P300-based lie detection," Comput. Meth. Programs Biomed., vol. 94, no. 1, pp. 48-57, 2009.

[10] M. Han, and X. Liu, "Feature selection techniques with class separability for multivariate time series," Neurocomputing, vol. 110, pp. 29-34, 2013.

[11] A. Gupta, R. K. Agrawal, and B. Kaur, "Performance enhancement of mental task classification using EEG signal: a study of multivariate feature selection methods," Soft Comput., vol. 19, no. 10, pp. 2799-2812, 2015.

[12] 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, 2005.

[13] M. Han, and W. Ren, "Global mutual information-based feature selection approach using single-objective and multi-objective optimization," Neurocomputing, vol. 168, pp. 47-54, 2015.

[14] J. Kennedy, and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," In International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation, Orlando, Florida, October 12-15, 1997, pp. 4104-4108.

[15] S. Mirjalili, and A. Lewis, "S-shaped versus V-shaped transfer functions for binary particle swarm optimization," Swarm Evol. Comput., vol. 9, pp. 1-14, 2013.

[16] E. D. Übeyli, "Wavelet/mixture of experts network structure for EEG signals classification," Expert Syst. Appl., vol. 34, no. 3, pp. 1954-1962, 2008.

[17] Y. Song, and P. Liò, "A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine," Journal of Biomedical Science and Engineering, vol. 3, no. 6, pp. 556-567, 2010.

[18] E. D. Übeyli, "Statistics over features: EEG signals analysis," Comput. Biol. Med., vol. 39, no. 8, pp. 733-741, 2009.

[19] N. F. Güler, E. D. Übeyli, and İ. Güler, "Recurrent neural networks employing Lyapunov exponents for EEG signals classification," Expert Syst. Appl., vol. 29, no. 3, pp. 506-514, 2005.

[20] U. R. Acharya, R. Yanti, G. Swapna, V. S. Sree, R. J. Martis, and J. S. Suri, "Automated diagnosis of epileptic electroencephalogram using independent component analysis and discrete wavelet transform for different electroencephalogram durations," Proc. Inst. Mech. Eng. Part H-J. Eng. Med., vol. 227, no. 3, pp. 234-244, 2013.

[21] 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, 2012.