Research Article

Sparse Representation-Based Extreme Learning Machine for Motor Imagery EEG Classification

Qingshan She [1], **Kang Chen,** [1] **Yuliang Ma** [1] **Thinh Nguyen** [2] **and Yingchun Zhang** [1,2]

[1]*Institute of Intelligent Control and Robotics, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China*
[2]*Department of Biomedical Engineering, University of Houston, Houston, TX 77204, USA*

Correspondence should be addressed to Qingshan She; qsshe@hdu.edu.cn and Yingchun Zhang; yzhang94@uh.edu

Classification of motor imagery (MI) electroencephalogram (EEG) plays a vital role in brain-computer interface (BCI) systems. Recent research has shown that nonlinear classification algorithms perform better than their linear counterparts, but most of them cannot extract sufficient significant information which leads to a less efficient classification. In this paper, we propose a novel approach called FDDL-ELM, which combines the discriminative power of extreme learning machine (ELM) with the reconstruction capability of sparse representation. Firstly, the common spatial pattern (CSP) algorithm is adopted to perform spatial filtering on raw EEG data to enhance the task-related neural activity. Secondly, the Fisher discrimination criterion is employed to learn a structured dictionary and obtain sparse coding coefficients from the filtered data, and these discriminative coefficients are then used to acquire the reconstructed feature representations. Finally, a nonlinear classifier ELM is used to identify these features in different MI tasks. The proposed method is evaluated on 2-class Datasets IVa and IIIa of BCI Competition III and 4-class Dataset IIa of BCI Competition IV. Experimental results show that our method achieved superior performance than the other existing algorithms and yielded the accuracies of 80.68%, 87.54%, and 63.76% across all subjects in the above-mentioned three datasets, respectively.

1. Introduction

The brain-computer interface (BCI) is a system that allows its users to use their brain activity to control external devices which are independent of peripheral nerves and muscles [1, 2]. Motor imagery- (MI-) based sensorimotor rhythm (SMR) analysis, including mu (8–14 Hz) and/or beta (15–30 Hz) rhythms, recorded from the scalp over the sensorimotor cortex, is one of the widely used methods in the BCI field [3, 4]. However, these MI signals are highly nonstationary and inevitably contaminated with noise, and meanwhile, they strongly depend on subjects [5].

Sparse representation (SR), originally proposed by Olshausen et al. [6], attempts to simulate the working mechanism of primary visual cortex in the human visual system. The basic idea is to represent the data as a linear combination of atoms in a dictionary, whose requirement is that the coefficients are sparse, i.e., they contain only a small number of nonzero elements. In the last two decades, SR has been widely studied for reconstruction, representation, and compression of high-dimensional noisy data, such as computer vision, pattern recognition, and bioinformatics [7–9]. Recently, the SR techniques have also yielded promising results in the BCI systems [10–15]. Although SR is a powerful tool to reconstruct the originals from noisy and imperfect data, using the original training samples as the dictionary may not fully exploit the discriminative information hidden in the training samples. To address the problem, Yang et al. [16] proposed a Fisher discrimination dictionary learning (FDDL) framework to learn a structured dictionary and had good reconstruction capability for the training samples, yielding a 3.2% improvement over the sparse representation-based classification (SRC) algorithm on AR datasets in face recognition.

Recently, Huang et al. developed a new efficient learning algorithm called extreme learning machine (ELM) [17, 18]

for training single-layer feedforward neural networks (SLFNs), featuring faster learning speed and better generalization capability in comparison with the well-known back propagation (BP) neural networks and support vector machines (SVMs). ELM has been applied to pattern recognition tasks in the BCI systems and has shown superior performance over traditional classification approaches [19–22]. In light of this advancement, efforts have been made in developing algorithms to integrate ELM and SR, thus exploiting the speed advantage and discriminative power of ELM and the antinoise performance and reconstruction ability of SR. A recent approach called extreme sparse learning (ESL) has been proposed in [23], which simultaneously learns sparse representation of the input signal and trains the ELM classifier. In the study by Yu et al. [24], the sparse coding technique is adopted to map the inputs to the hidden layer, instead of the random mapping used in classic ELM. Other ELM-SR hybrid models were also extensively studied, in which the ELM classifier is firstly employed to estimate noisy signals, and then a further identification for the estimated signals is carried out using the SRC algorithm [25–27].

Most of the existing ELM methods employ a single hidden layer network structure. While benefitting from a relatively fast training speed, it is well known that for a single hidden layer network, the training sample is always the original training sample set, which could limit the robustness of the network. Furthermore, due to its shallow architecture, feature learning using SLFNs may not be effective for natural signals (e.g., EEG). To incorporate a deeper network structure, a hierarchical-extreme learning machine (H-ELM) method has recently been developed, which allows for a layer-wise architecture design, and have shown to yield great classification performance [28]. In addition, multilayer structure has been extended into ELM in [29, 30] as well. Inspired by these works, we propose a new layer-wise structure framework called FDDL-ELM, which combines the idea of SR with ELM to learn a powerful nonlinear classifier. The proposed method first employs the Fisher discrimination criterion to learn a structured dictionary. With the learned dictionary, more discriminative sparse coding coefficients can be obtained, and more robust feature information can be extracted. Subsequently, the ELM classifier is utilized to discriminate the extracted features. The classification accuracy of the proposed method has been manifested by several benchmark datasets, as well as 2-class and 4-class real world EEG data from BCI Competition III Datasets IVa and IIIa and BCI Competition IV Dataset IIa.

The rest of the paper is organized as follows: Section 2 presents a brief introduction to basic ELM and FDDL and provides detailed description of the proposed FDDL-ELM algorithm. Section 3 evaluates the performance of the FDDL-ELM method through a series of experiments on several benchmark datasets, as well as motor imagery EEG datasets. Finally, we will conclude the paper and present some future work in Section 4.

2. Methodology

2.1. Classic ELM.

ELM was originally implemented for single-hidden layer feedforward neural networks and then extended to generalize feedforward networks. By using random hidden node parameters and tuning-free strategy, ELM has some notable advantages, such as easy implementation, fast learning speed, and superior generalization performance [17], thus making it a suitable choice for the recognition problem of EEG signals in different motor imagery tasks.

Consider a dataset containing N training samples, $\{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_i, \mathbf{y}_i\}$, $i = 1, 2, \cdots, N$, with the input $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]^{\mathrm{T}} \in R^p$ and its corresponding desired output $\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{iq}]^{\mathrm{T}} \in R^q$, where T denotes a transpose operation. Assuming that m is the number of hidden neurons, and $g(\cdot)$ is the activation function, the output function of ELM is mathematically modeled as

$$\mathbf{y}_j = \sum_{i=1}^{m} \boldsymbol{\beta}_i g\left(\mathbf{a}_i^{\mathrm{T}} \mathbf{x}_j + b_i\right), \quad j = 1, 2, \cdots, N, \tag{1}$$

where $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{iq}]^{\mathrm{T}}$ is the weight vector that connects the i-th hidden neuron and the output neurons, $\mathbf{a}_i = [a_{i1}, a_{i2}, \cdots, a_{ip}]^{\mathrm{T}}$ is the randomly chosen input weight vector connecting the i-th hidden neuron and the input neurons, b_i is the randomly chosen bias of the i-th hidden node, and y_j is the actual output corresponding to input x_j.

For convenience of expression, the Equation (1) is written in matrix notation as

$$\mathbf{Y} = \mathbf{H}\boldsymbol{\beta}, \tag{2}$$

where $\mathbf{Y} = [y_1, y_2, \cdots, y_N]_{N \times q}^{\mathrm{T}}$ is the expected network output, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m]_{m \times q}^{\mathrm{T}}$ denotes the weight of output layer, and $\mathbf{H}$ is the hidden layer output matrix which is defined as

$$\mathbf{H} = \begin{bmatrix} g\left(\mathbf{a}_1^{\mathrm{T}}\mathbf{x}_1 + b_1\right) & \ldots & g\left(\mathbf{a}_m^{\mathrm{T}}\mathbf{x}_1 + b_m\right) \\ \ldots & \ldots & \ldots \\ g\left(\mathbf{a}_1^{\mathrm{T}}\mathbf{x}_N + b_1\right) & \ldots & g\left(\mathbf{a}_m^{\mathrm{T}}\mathbf{x}_N + b_m\right) \end{bmatrix}_{N \times m}. \tag{3}$$

To have better generalization performance, the regularization parameter C is introduced in [19], and its corresponding objective function is given by

$$\underset{\boldsymbol{\beta}}{\arg\min}\left(\|\boldsymbol{\beta}\|_2^2 + C\|\mathbf{H}\boldsymbol{\beta} - \mathbf{Y}\|_2^2\right), \tag{4}$$

where $\|\cdot\|_2$ denotes the l_2-norm of a matrix or a vector. We can obtain the output weight vector $\boldsymbol{\beta}$ using the Moore–Penrose principle. The solution of Equation (4) is $\boldsymbol{\beta} = (\mathbf{I}/C + \mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{Y}$ if $N > m$ and $\boldsymbol{\beta} = \mathbf{H}^{\mathrm{T}}(\mathbf{I}/C + \mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{Y}$ if $N < m$.

2.2. Fisher Discrimination Dictionary Learning.

Sparse representation-based classification (SRC) was proposed for face recognition, which directly used the training samples of all classes as the dictionary to code the query face image and

classified it by evaluating which class leads to the minimal reconstruction error [7]. However, the dictionary in use may not be effective enough to represent the query images due to the uncertain and noisy information in the original training images, and the discriminative information hidden in the training samples is not sufficiently exploited by such a naïve-supervised dictionary learning approach [16]. To address these problems, the FDDL method is proposed, utilizing both the discriminative information in the reconstruction error and sparse coding coefficients.

Denote $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_c]$ as the training set, where $\mathbf{A}_i$ is the subset of the training samples from class i, and c is the total number of classes, and an overcomplete dictionary $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_c]$, where $\mathbf{D}_i$ is the class-specified sub-dictionary associated with class i. Let $\mathbf{X}$ be the coding coefficient matrix of $\mathbf{A}$ over $\mathbf{D}$, we can write $\mathbf{X}$ as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_c]$, where $\mathbf{X}_i$ is the submatrix containing the coding coefficients of $\mathbf{A}_i$ over $\mathbf{D}$. The objective function is written as follows:

$$J_{(\mathbf{D},\mathbf{X})} = \arg \min_{\mathbf{D},\mathbf{X}} \{ r(\mathbf{A}, \mathbf{D}, \mathbf{X}) + \lambda_1 \|\mathbf{X}\|_1 + \lambda_2 f(\mathbf{X}) \}, \quad (5)$$

where $r(\mathbf{A}, \mathbf{D}, \mathbf{X})$ is the discriminative fidelity term, $\|\mathbf{X}\|_1$ is the sparse constraint in which the notation $\|\cdot\|_1$ denotes the l_1-norm, $f(\mathbf{X})$ is a discrimination constraint, and λ_1 and λ_2 are the scalar parameters.

$$r(\mathbf{A}_i, \mathbf{D}, \mathbf{X}_i) = \|\mathbf{A}_i - \mathbf{D}\mathbf{X}_i\|_F^2 + \|\mathbf{A}_i - \mathbf{D}_i\mathbf{X}_i^i\|_F^2 + \sum_{j=1, j \neq i}^{c} \|\mathbf{D}_j\mathbf{X}_i^j\|_F^2, \quad (6)$$

where $\|\cdot\|_F$ means the F-norm, $\mathbf{X}_i^i$ is the coding coefficient of $\mathbf{A}_i$ over the subdictionary $\mathbf{D}_i$, and $\mathbf{X}_i^j$ is the coding coefficient of $\mathbf{A}_i$ over the subdictionary $\mathbf{D}_j$. The minimization of $r(\mathbf{A}, \mathbf{D}, \mathbf{X})$ means that the reconstruction error of the i-th class of samples is minimized, and the reconstruction error through the i-th subdictionary is also minimized while the reconstruction by other subclass dictionaries should be minimized. Its purpose is to ensure that the reconstruction error constraint is minimized, and the sparse coefficient can be more discriminative.

$f(\mathbf{X})$ is a discriminative coefficient term which is given in the following:

$$f(\mathbf{X}) = \text{tr}(S_W(\mathbf{X})) - \text{tr}(S_B(\mathbf{X})) + \eta \|\mathbf{X}\|_F^2, \quad (7)$$

where $\text{tr}(\cdot)$ indicates the trace of subspace, $S_W(\mathbf{X})$ is the within-class scatter of $\mathbf{X}$, $S_B(\mathbf{X})$ is the between-class scatter of $\mathbf{X}$, and η is a parameter.

$$S_W(\mathbf{X}) = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in \mathbf{X}_i} (\mathbf{x}_k - \mathbf{m}_i)(\mathbf{x}_k - \mathbf{m}_i)^T,$$
$$S_B(\mathbf{X}) = \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T, \quad (8)$$

where $\mathbf{m}_i$ and $\mathbf{m}$ are the mean vectors of $\mathbf{X}_i$ and $\mathbf{X}$ respectively, and n_i is the number of samples in class $\mathbf{A}_i$.

2.3. The Proposed FDDL-ELM Method.

In this section, we propose a novel nonlinear classification model that rests on a new ELM framework for multilayer perceptron (MLP), named FDDL-ELM. The framework consists of two stages: an encoding stage and a classification stage. The former stage uses the FDDL approach to map the input features into a midlevel feature space, and then the ELM algorithm is performed for final decision making in the latter stage. The framework of FDDL-ELM is shown in Figure 1.

Let $\mathbf{A}$ be an input containing N training samples $\{\mathbf{A}, \mathbf{Y}\}$ with $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_c]$, and $\mathbf{Y}$ is the corresponding desired output, where $\mathbf{A}_i$ is the subset of the input from class i, and c is the total number of classes.

Step (1): utilize the FDDL algorithm to learn a structured dictionary $\mathbf{D}$.

By incorporating Equations (6) and (7) into Equation (5), the objective function is rewritten as

$$J_{(\mathbf{D},\mathbf{X})} = \arg \min_{(\mathbf{D},\mathbf{X})} \left\{ \sum_{i=1}^{c} r(\mathbf{A}_i, \mathbf{D}, \mathbf{X}_i) + \lambda_1 \|\mathbf{X}\|_1 + \lambda_2 (\text{tr}(S_W(\mathbf{X}) - S_B(\mathbf{X})) + \eta \|\mathbf{X}\|_F^2) \right\}. \quad (9)$$

The optimization of the objective function consists of two steps: First, update $\mathbf{X}$ by fixing $\mathbf{D}$, and then update $\mathbf{D}$ while fixing $\mathbf{X}$. The procedures are iteratively implemented for the desired discriminative dictionary $\mathbf{D}$ and the discriminative coefficients $\mathbf{X}$ as done in [16].

Step (2): reconstruct the signals for the high-level sparse feature information.

With the desired dictionary $\mathbf{D}$ and the coefficients $\mathbf{X}$ in the Step (1), we can get the reconstructed signals $\mathbf{B}$ which can uncover important information hidden in the original signals and is simplified as follows:

$$\mathbf{B} = \mathbf{D}\mathbf{X}. \quad (10)$$

Step (3): discriminate the reconstructed signals $\mathbf{B}$ using the ELM classification method.

(1) Randomly generate the hidden node parameters $(\mathbf{a}_i, b_i)$ for $i = 1, 2, \cdots, m$.

(2) The new hidden-layer output matrix $\mathbf{G}$ can be written as

$$\mathbf{G} = \begin{bmatrix} g(\mathbf{a}_1^T \mathbf{D}\mathbf{X}_1 + b_1) & \cdots & g(\mathbf{a}_m^T \mathbf{D}\mathbf{X}_1 + b_m) \\ \cdots & \cdots & \cdots \\ g(\mathbf{a}_1^T \mathbf{D}\mathbf{X}_N + b_1) & \cdots & g(\mathbf{a}_m^T \mathbf{D}\mathbf{X}_N + b_m) \end{bmatrix}_{N \times m}. \quad (11)$$

(3) The regularization parameter C is introduced, and the output weight $\boldsymbol{\beta}$ is calculated as follows:

$$\boldsymbol{\beta} = \left(\frac{\mathbf{I}}{C} + \mathbf{G}^T\mathbf{G}\right)^{-1} \mathbf{G}^T\mathbf{Y}, \quad \text{for } N > m,$$
$$\boldsymbol{\beta} = \mathbf{G}^T \left(\frac{\mathbf{I}}{C} + \mathbf{G}\mathbf{G}^T\right)^{-1} \mathbf{Y}, \quad \text{for } N < m. \quad (12)$$
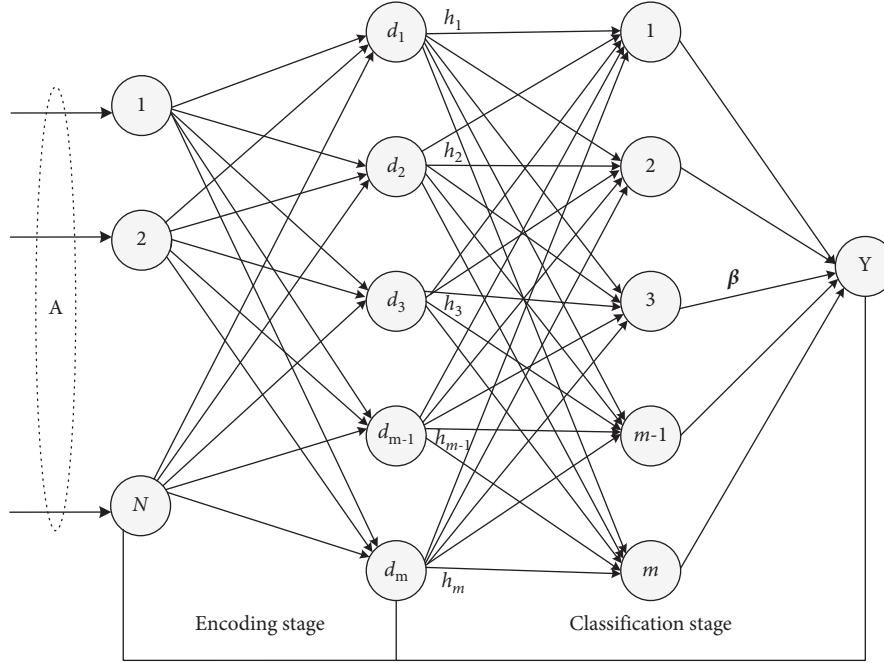
Figure 1: A schematic for the overall framework of the FDDL-ELM-learning algorithm.

Extensive efforts have been paid to the optimal selection of C and the leave-one-out (LOO) cross-validation strategy combining with the predicted residual sum of squares (PRESS) statistic is one of the most effective methods [26].

Step (4): For test data $\{\mathbf{A}_{\text{test}}, \mathbf{Y}_{\text{test}}\}$ and the learned dictionary $\mathbf{D}$, we can reconstruct the $\mathbf{A}_{\text{test}}$ in the encoding stage and then calculate the labels $\mathbf{Y}_{\text{predict}}$ using the ELM classifier.

3. Experimental Results and Discussion

In this section, several experiments on benchmark datasets and EEG datasets were performed to evaluate the performance of the proposed FDDL-ELM method, as compared with the other state-of-the-art approaches. All methods were implemented using MATLAB 2014b environment on a computer with a 2.6 GHz processor and 8.0 GB RAM.

3.1. Experiment on Benchmark Datasets

3.1.1. Description. In order to evaluate its performance, the proposed FDDL-ELM method was first applied to four popular benchmark datasets in the UCI repository [31]. The details of these datasets are shown in Table 1.

The Liver Disorders dataset is a medical application, which consists of 345 samples belonging to 2 categories, and each sample extracts 6 features for representation. The Diabetes dataset contains 768 samples belonging to two categories. For each sample, 8 features are extracted. The Waveform dataset consists of 5000 samples from 3 classes of noisy waveforms, and each sample contains 21 attributes. The Columbia Object Image Library (COIL-20) is a multi-class image classification dataset and consists of 1440 grayscale image sample of 20 different objects, in which each

Table 1: Description of the benchmark datasets.

Datasets	Training	Testing	Features	Classes	Random perm
Liver Disorders	172	172	6	2	Yes
Diabetes	384	384	8	2	Yes
Waveform	2500	2500	21	3	Yes
COIL-20	720	720	1024	20	Yes

sample is a 32×32 grayscale image of one object taken from a specific view.

3.1.2. Experimental Setup. In Table 1, the column "Random perm" denotes whether the training and test data are randomly assigned. In each data partition, the ratio between training and test sample is 1 : 1. The classification process was repeated ten times, and the average of these outcomes was the final classification rate.

This proposed FDDL-ELM algorithm has 5 tuning parameters: λ_1, λ_2, and η in the encoding stage, as well as the number of hidden nodes m, and the regularization parameter C in the classification stage. In all the experiments, the optimal parameters λ_1 and λ_2 are searched using five-fold cross-validation from a small set $\{0.001, 0.005, 0.01, 0.05, 0.1\}$, and η is set to 1, as done in [16]. The optimal parameters m and C were determined from $m \in \{100, 200, \cdots, 1500\}$ and $C \in \{e^{-5}, e^{-4}, \cdots, e^5\}$ using the LOO cross-validation strategy based on the minimum MSE^{PRES} [27]. It is noted that C is automatically chosen and not fixed during the process of repeating ten times in the classification stage. The settings of these tuning parameters for four benchmark datasets are summarized in Table 2.

TABLE 2: Parameter settings of FDDL-ELM on the benchmark datasets.

Datasets	λ_1	λ_2	m
Liver Disorders	0.01	0.001	400
Diabetes	0.005	0.05	500
Waveform	0.001	0.01	100
COIL-20	0.05	0.001	900

TABLE 3: Comparisons of classification results on each dataset using different methods.

Method	Liver Disorders Acc ± sd	Diabetes Acc ± sd	Waveform Acc ± sd	COIL-20 Acc ± sd
ELM	72.15 ± 1.54	74.22 ± 1.10	84.45 ± 0.69	96.14 ± 1.03
FDDL	65.77 ± 4.45	65.43 ± 2.49	79.18 ± 1.19	96.48 ± 0.85
H-ELM	**74.01 ± 0.87**	71.67 ± 1.26	84.72 ± 0.30	97.13 ± 0.71
FDDL-ELM	72.38 ± 1.49	**75.33 ± 0.68**	**85.02 ± 0.32**	**98.33 ± 0.62**

3.1.3. Comparisons with Other State-of-the-Art Algorithms. In this experiment, we compare the proposed FDDL-ELM with three baseline algorithms, including ELM, FDDL, and H-ELM. The classification performance is evaluated in terms of average accuracy and standard deviation (acc ± sd). Table 3 summarizes the performance results using four methods on the benchmark datasets.

From the results shown in Table 3, it is evidenced that the FDDL-ELM algorithm achieved comparable performance with other state-of-the-art methods, such as single-layer ELM, FDDL, and H-ELM with deep architecture. For the Diabetes dataset, the FDDL-ELM approach achieved more than 9% improvement over FDDL. For the Liver Disorders dataset, although the average classification accuracy of H-ELM (74.01%) was better than that of FDDL-ELM (72.38%), the accuracy of FDDL-ELM was higher than ELM by 0.23% and FDDL by 6.61%. For the Waveform dataset, the FDDL-ELM approach yielded a mean accuracy of 85.02%, a 0.57% improvement over ELM, and a 0.30% improvement over H-ELM. The average classification accuracy of COIL-20 dataset obtained by FDDL-ELM was 98.33%, higher than those of ELM (96.14%) and H-ELM (97.13%). Based on these observations, the proposed FDDL-ELM approach outperformed the original ELM and FDDL methods on all four datasets and had comparable performance on most of the four datasets compared with H-ELM.

3.1.4. The Impact of the Parameters. There are five parameters in our algorithm: λ_1, λ_2, η, m, and C. Since η is set to 1 [16], and C is automatically chosen [27], we will investigate the impact of the other three parameters (λ_1, λ_2, and m) on the performance of our algorithm in this section. λ_1 and λ_2 are respectively changed among $\{0.001, 0.005, 0.01, 0.05, 0.1\}$. The parameter m decides the number of hidden neurons, and its value is selected among $\{100, 200, \cdots, 1500\}$.

Figures 2(a)–5(a) show the testing results of our algorithm as the parameter m changes on four datasets: Diabetes, Liver Disorders, Waveform, COIL-20 datasets, respectively. As can be seen, the performance of our algorithm is relatively stable with respect to m. Figures 2(b)–5(b) give the plots of testing accuracies as λ_1 and λ_2 vary on four datasets, respectively. From these results, it can be observed that the performance is more pronouncedly affected by the parameters λ_1 and λ_2, relative to m. For a tradeoff between the classification performance and computation complexity, we can use a comparatively small number of nodes m when selecting λ_1 and λ_2 in real EEG applications.

3.2. Experiment on BCI Datasets

3.2.1. Description. This section evaluates the performance of the proposed FDDL-ELM method on MI EEG datasets. There are three datasets for analysis, including two datasets for binary classification and one dataset for multi-classification, as described below:

(1) Dataset IVa, BCI competition III [32]: this dataset contains EEG signals from 5 subjects, who performed 2-class MI tasks: right hand and foot. EEG signals were recorded using 118 electrodes. A training set and a testing set were available for each subject. Their size was different for each subject. More precisely, 280 trials were available for each subject, among which 168, 224, 84, 56, and 28 composed the training set for subjects A1, A2, A3, A4, and A5, respectively, and the remaining trials composed the testing set.

(2) Dataset IIIa, BCI competition III [33]: this dataset comprised EEG signals from 3 subjects who performed left hand, right hand, foot, and tongue MI. EEG signals were recorded using 60 electrodes. For the purpose of binary classification, only 2-class EEG signals (left and right hand MI) were used as done in [34]. Both the training and testing sets were available for each subject. Both sets contain 45 trials per class for subject B1, and 30 trials per class for subjects B2 and B3.

(3) Dataset IIa, BCI competition IV [35]: this dataset consists of EEG signals from 9 subjects who performed 4-class MI tasks: left hand, right hand, foot, and tongue MI. EEG signals were recorded using 22 electrodes. The training and testing sets contain 288 trials for each class, respectively.

3.2.2. Experimental Setup. Data preprocessing was first performed on the raw EEG data. In particular, for each trial, we extracted features from the time-segment spanning from 0.5 s to 2.5 s after the cue instructing the subject to perform MI. Each trial was band-pass filtered in 8–30 Hz, using a fifth-order Butterworth filter. Next, the dimension of the EEG signal was reduced using the common spatial pattern (CSP) algorithm, a widely used feature selection method for MI-based BCIs [12, 34]. Finally, the filtered EEG signals by CSP were discriminated by different classification methods in our experiment.
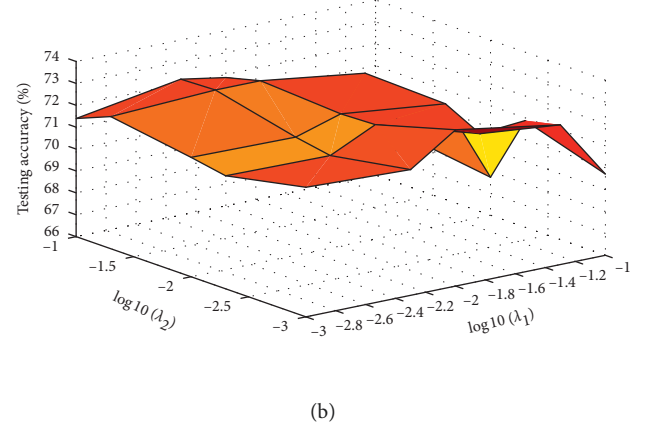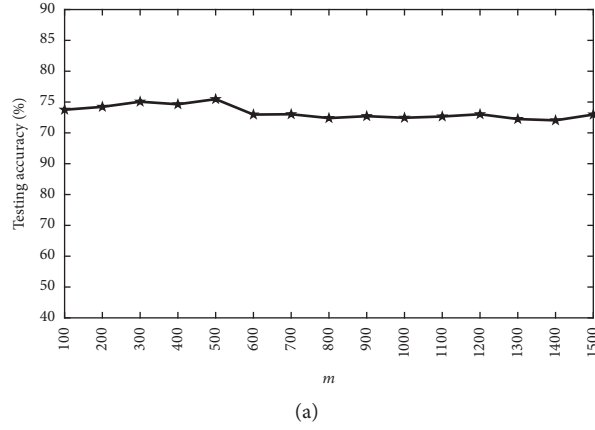
(a)

(b)

FIGURE 2: Testing accuracy with different parameters on Diabetes. (a) Accuracy in terms of m; (b) accuracy curve in terms of (λ_1, λ_2).
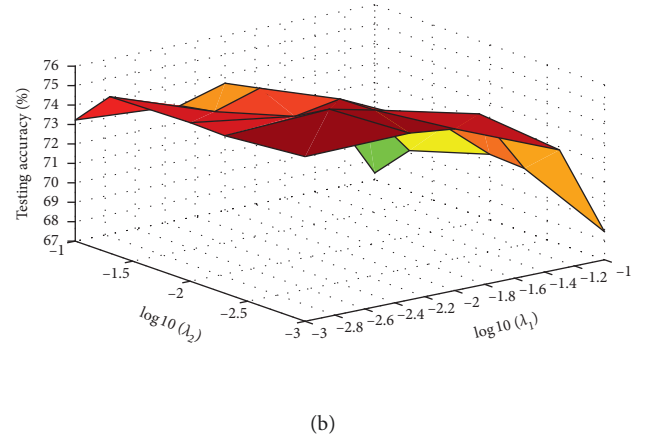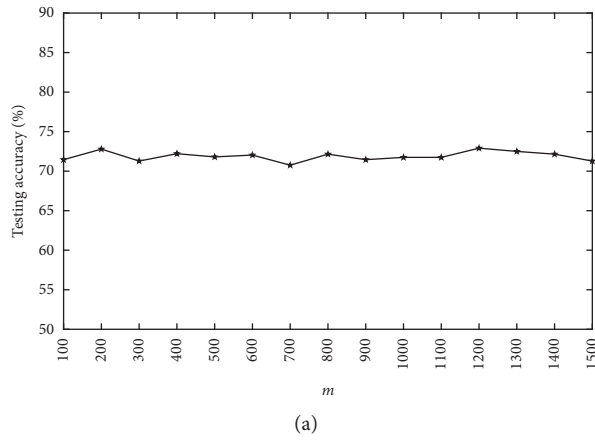

(a)

(b)

FIGURE 3: Testing accuracy with different parameters on liver disorders. (a) Accuracy in terms of m; (b) accuracy curve in terms of (λ_1, λ_2).
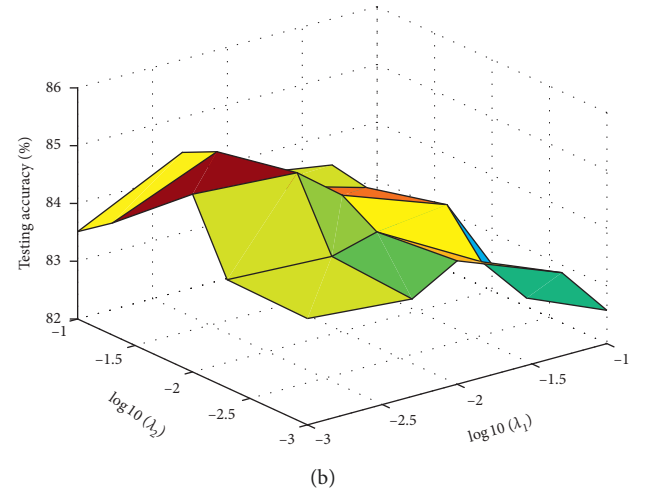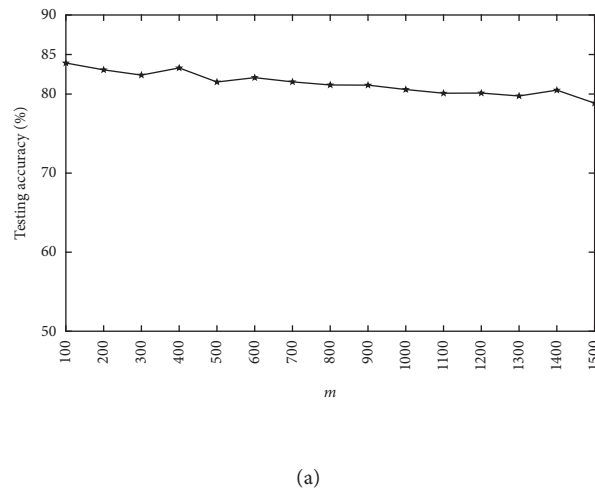

(a)

(b)

FIGURE 4: Testing accuracy with different parameters on waveform. (a) Accuracy in terms of m; (b) accuracy curve in terms of (λ_1, λ_2).
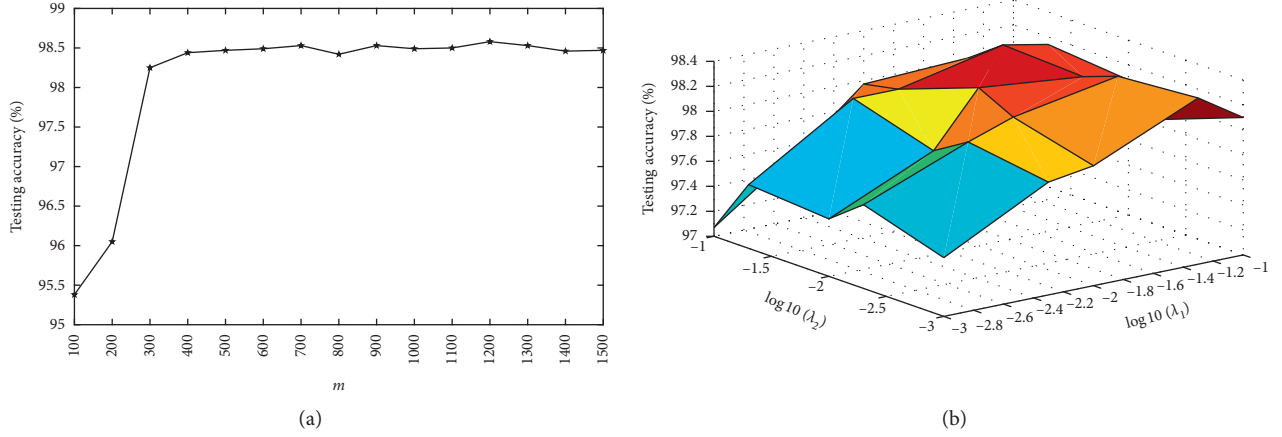
FIGURE 5: Testing accuracy with different parameters on COIL-20. (a) Accuracy in terms of m; (b) accuracy curve in terms of (λ_1, λ_2).

In this work, the classification process was repeated ten times, and the average accuracy was recorded for further analysis. The selection process of the parameters λ_1, λ_2, η, and C were the same as those described in Section 3.1.2, and the setting of the hidden node m was $\{10, 20, \cdots, 100\}$.

3.2.3. Comparisons with Related Algorithms.

We compared the proposed method FDDL-ELM with ELM, FDDL, and H-ELM on BCI Competition III Datasets IVa and IIIa and BCI Competition IV Dataset IIa. The average classification accuracies of all four algorithms are shown in Table 4.

From Table 4, it can be seen that the proposed method outperformed the ELM and FDDL algorithms on almost all subjects (except subject B2) in binary-classification applications. For subject B2, the ELM method obtained the average accuracy of 68.33%, a 0.33% improvement over FDDL-ELM. Compared with H-ELM which adopts a deep architecture, FDDL-ELM yielded comparable performance on all the 8 subjects and especially performed better in 4 subjects (A2, A3, A4, and B2). Furthermore, the proposed algorithm yielded the highest average accuracy on Datasets IVa and IIIa. For the Dataset IVa, the FDDL-ELM approach achieved a mean accuracy of 80.68%, a 0.73% improvement over ELM, and a 1.35% improvement over H-ELM. Moreover, a paired t-test revealed no significant difference between the FDDL-ELM and H-ELM approaches ($p = 0.626$) and a significant difference between the FDDL-ELM and ELM approaches ($p = 0.04$). For the Dataset IIIa, the average classification accuracy obtained by FDDL-ELM was 87.54%, higher than that of ELM (87.35%), FDDL (83.26%), and H-ELM (85.63%). Furthermore, a paired t-test revealed no significant difference between the FDDL-ELM and H-ELM approaches ($p = 0.596$). These results have shown that the FDDL-ELM method has achieved a great classification capacity in binary-classification applications.

In the 4-class-classification application for BCI Competition IV Dataset IIa, the average classification accuracies for the 9 subjects using four algorithms are also shown in Table 4. Note that our method also outperformed ELM and FDDL in 8 of the 9 subjects (except subject C8). For subject C8, ELM gained the best result (81.87%) compared with

TABLE 4: Comparisons of classification results on 2-class and 4-class BCI datasets using different methods.

Datasets		Methods			
		ELM	FDDL	FDDL-ELM	H-ELM
Dataset IVa	A1	60.71	57.50	61.70	**63.39**
	A2	100	84.29	**100**	98.39
	A3	73.37	70.51	**73.88**	64.08
	A4	86.61	65.00	**88.17**	85.67
	A5	79.05	77.14	79.64	**85.16**
	Mean	79.95	70.89	**80.68**	79.33
Dataset IIIa	B1	96.89	93.11	97.78	**98.56**
	B2	**68.33**	60.33	68.00	60.00
	B3	96.83	96.33	96.83	**98.33**
	Mean	87.35	83.26	**87.54**	85.63
Dataset IIa	C1	76.43	62.77	**76.74**	75.69
	C2	45.38	32.92	45.70	**47.74**
	C3	76.32	70.76	**77.13**	76.98
	C4	59.24	45.59	60.50	**61.84**
	C5	37.12	31.42	36.24	**37.85**
	C6	45.90	35.28	**47.57**	47.08
	C7	78.99	66.91	**80.30**	80.07
	C8	**81.87**	63.72	80.60	76.46
	C9	67.36	65.48	69.10	**76.42**
	Mean	63.18	52.76	63.76	**64.46**

FDDL (63.72%), FDDL-ELM (80.60%), and H-ELM (76.46%). The FDDL-ELM approach performed the best in 4 subjects (C1, C3, C6, and C7), whereas H-ELM achieved the best result in 4 subjects (C2, C4, C5, and C9). The average classification accuracy of 9 subjects using FDDL-ELM was 63.76%, slightly lower than H-ELM (64.46%). A paired t-test revealed no significant difference between the FDDL-ELM and H-ELM approaches ($p = 0.519$) and a significant difference between the FDDL-ELM and FDDL approaches (p is less than 0.01). These results showed that when compared to the H-ELM algorithm, our method can achieve similar results without the deep architecture.

In these experiments, the proposed FDDL-ELM method exhibited an excellent performance in both binary-classification and multiclassification cases. The nonlinear property of FDDL-ELM allowed for its superior performance over the FDDL approach when processing the

nonstationary EEG signals. Furthermore, FDDL-ELM is more suitable in analyzing noisy EEG data than basic ELM because its encoding stage can acquire a higher representation of the raw signals and extract more effective feature information. Compared with the H-ELM, an algorithm with a deep architecture design, our method also yielded comparable results. In particular, on the binary-classification datasets (BCI Competition III Datasets IVa and IIIa), our method gained higher average accuracies (80.68% and 87.54%) than that of H-ELM (79.33% and 85.63%), respectively.

4. Conclusion

In this paper, we have proposed a new ELM framework called FDDL-ELM, which achieves a sparse representation of input raw data with layer-wise encoding, while still benefiting from the universal approximation capability of the original ELM. We verified the generalizability and capability of FDDL-ELM using publicly available benchmark databases and MI-BCI datasets. In these applications, FDDL-ELM demonstrated superior classification performance than the other relevant state-of-the-art methods. However, there are still several questions to be further investigated in future work. The nonstationary nature of EEG signals means that a classification model built earlier using the previous data is not able to well reflect the changes that have already taken place to the signals. Consequently, the online updates to the classification model are needed. Recently, the ensemble of subset online sequential extreme learning machine (ESOS-ELM) method is proposed for class imbalance learning [36]. In addition, an online sequential extreme learning machine with kernels (OS-ELMK) has been proposed for prediction of nonstationary time series [37]. In the study by Mirza et al.[38], a multilayer online sequential extreme learning machine has been proposed for image classification. In the future work, we will investigate the online learning algorithm of FDDL-ELM for analyzing MI EEG signals.

Data Availability

Three datasets were employed in this study, including two datasets for binary classification and one dataset for multiclassification, which are publicly available: (1) dataset IVa, BCI competition III [31]: this dataset contains EEG signals from 5 subjects, who performed 2-class MI tasks: right hand and foot. (2) Dataset IIIa, BCI competition III [34]: this dataset comprised EEG signals from 3 subjects who performed left hand, right hand, foot, and tongue MI. (3) Dataset IIa, BCI competition IV [32]: this dataset consists of EEG signals from 9 subjects who performed 4-class MI tasks: left hand, right hand, foot, and tongue MI.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

References

[1] S. Sun and J. Zhou, "A review of adaptive feature extraction and classification methods for EEG-based brain-computer interfaces," in *Proceedings of IEEE International Joint Conference on Neural Networks (IJCNN)*, pp. 1746–1753, Beijing, China, 2014.

[2] J. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*, OUP, Oxford, UK, 2012.

[3] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selection in motor imagery based brain-computer interface," in *Proceedings of IEEE International Conference on Engineering in Medicine and Biology Society*, vol. 5, pp. 5392–5395, Shanghai, China, 2005.

[4] Q. She, H. Gan, Y. Ma et al., "Scale-dependent signal identification in low-dimensional subspace: motor imagery task classification," *Neural Plasticity*, vol. 2016, pp. 1–15, Article ID 7431012, 2016.

[5] C. Park, D. Looney, N. ur Rehman et al., "Classification of motor imagery BCI using multivariate empirical mode decomposition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 1, pp. 10–22, 2013.

[6] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1," *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[7] J. Wright, A. Y. Yang, A. Ganesh et al., "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[8] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, 2014.

[9] Y. A. Huang, Z. H. You, X. Chen et al., "Sequence-based prediction of protein-protein interactions using weighted sparse representation model combined with global encoding," *BMC Bioinformatics*, vol. 17, no. 1, pp. 1–11, 2016.

[10] Y. Shin, S. Lee, M. Ahn et al., "Noise robustness analysis of sparse representation based classification method for nonstationary EEG signal classification," *Biomedical Signal Processing and Control*, vol. 21, pp. 8–18, 2015.

[11] Q. Yuan, W. Zhou, S. Yuan et al., "Epileptic EEG classification based on kernel sparse representation," *International Journal of Neural Systems*, vol. 24, no. 04, Article ID 1450015, 2014.

[12] D. Wen, P. Jia, Q. Lian et al., "Review of sparse representation-based classification methods on EEG signal processing for Epilepsy detection, brain-computer interface and cognitive impairment," *Frontiers in Aging Neuroscience*, vol. 8, p. 172, 2016.

[13] W. Zhou, Y. Yang, and Z. Yu, "Discriminative dictionary learning for EEG signal classification in brain-computer interface," in *Proceedings of IEEE Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 1582–1585, Guangzhou, China, 2012.

[14] Y. Li, Z. L. Yu, N. Bi et al., "Sparse representation for brain signal processing: a tutorial on methods and applications," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 96–106, 2014.

[15] Y. Shin, S. Lee, J. Lee et al., "Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems," *Journal of Neural Engineering*, vol. 9, no. 5, Article ID 056002, 2012.

[16] M. Yang, L. Zhang, X. Feng et al., "Sparse representation based fisher discrimination dictionary learning for image classification," *International Journal of Computer Vision*, vol. 109, no. 3, pp. 209–232, 2014.

[17] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.

[18] A. Lendasse, Q. He, Y. Miche et al., "Advances in extreme learning machines (ELM2012)," *Neurocomputing*, vol. 128, pp. 1–3, 2014.

[19] L. Duan, M. Bao, J. Miao et al., "Classification based on multilayer extreme learning machine for motor imagery task from EEG signals," *Procedia Computer Science*, vol. 88, pp. 176–184, 2016.

[20] S. Ding, N. Zhang, X. Xu et al., "Deep extreme learning machine and its application in EEG classification," *Mathematical Problems in Engineering*, vol. 2015, no. 1, pp. 1–11, 2015.

[21] Y. Peng and B. L. Lu, "Discriminative manifold extreme learning machine and applications to image and EEG signal classification," *Neurocomputing*, vol. 174, pp. 265–277, 2016.

[22] Y. Zhang, J. Jin, X. Wang et al., "Motor imagery EEG classification via Bayesian extreme learning machine," in *Proceedings of IEEE International Conference on Information Science and Technology (ICIST)*, pp. 27–30, Guangzhou, China, 2016.

[23] S. Shojaeilangari, W. Y. Yau, K. Nandakumar et al., "Robust representation and recognition of facial emotions using extreme sparse learning," *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2140–2152, 2015.

[24] Y. Yu and Z. Sun, "Sparse coding extreme learning machine for classification," *Neurocomputing*, vol. 261, no. SI, pp. 50–56, 2017.

[25] M. Luo and K. Zhang, "A hybrid approach combining extreme learning machine and sparse representation for image classification," *Engineering Applications of Artificial Intelligence*, vol. 27, pp. 228–235, 2014.

[26] J. Cao, J. Hao, X. Lai et al., "Ensemble extreme learning machine and sparse representation classification," *Journal of the Franklin Institute*, vol. 353, no. 17, pp. 4526–4541, 2016.

[27] J. Cao, K. Zhang, M. Luo et al., "Extreme learning machine and adaptive sparse representation for image classification," *Neural networks*, vol. 81, pp. 91–102, 2016.

[28] J. Tang, C. Deng, and G. B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 4, pp. 809–821, 2016.

[29] T. V. Nguyen and B. Mirza, "Dual-layer kernel extreme learning machine for action recognition," *Neurocomputing*, vol. 260, pp. 123–130, 2017.

[30] B. Mirza, S. Kok, Z. Lin et al., "Efficient representation learning for high-dimensional imbalance data," in *Proceedings of 2016 IEEE Conference on Digital Signal Processing*, pp. 511–515, Beijing, China, 2017.

[31] C. C. Blake and C. J. Merz, "UCI repository of machine learning databases," http://archive.ics.uci.edu/ml/.

[32] Graz University, "BCI competition III datasets Iva," http://www.bbci.de/competition/iii/#datasetIva.

[33] Graz University, "BCI Competition III Datasets IIIa," http://www.bbci.de/competition/iii/#datasetIIIa.

[34] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[35] Graz University, "BCI competition IV datasets 2a," http://www.bbci.de/competition/iv/#dataset2a.

[36] B. Mirza, Z. Lin, and N. Liu, "Ensemble of subset online sequential extreme learning machine for class imbalance and concept drift," *Neurocomputing*, vol. 149, pp. 316–329, 2015.

[37] X. Wang and M. Han, "Online sequential extreme learning machine with kernels for nonstationary time series prediction," *Neurocomputing*, vol. 145, pp. 90–97, 2014.

[38] B. Mirza, S. Kok, and F. Dong, "Multi-layer online sequential extreme learning machine for image classification," in *Proceedings in Adaptation, Learning and Optimization*, vol. 1, pp. 39–49, Springer, New York, NY, USA, 2016.