IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 12, 2020, accepted June 17, 2020, date of publication June 22, 2020, date of current version July 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003916

A Novel Broad Learning Model-Based Semi-Supervised Image Classification Method

JIANJIE ZHENG[1], (Member, IEEE), YU YUAN[2], (Member, IEEE), HUIMIN ZHAO[3,4], AND WU DENG[1,3,5], (Member, IEEE)

[1]College of Software Engineering, Dalian Jiaotong University, Dalian 116028, China
[2]School of Locomotive and Rolling Stock Engineering, Dalian Jiaotong University, Dalian 116028, China
[3]College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China
[4]State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing 400044, China
[5]Traction Power State Key Laboratory, Southwest Jiaotong University, Chengdu 610031, China

Corresponding author: Wu Deng (dw7689@djtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771087, Grant 51605068, and Grant 51475065, in part by the Natural Science Foundation of Liaoning Province under Grant 2019ZD0099 and Grant 2019ZD0112, in part by the Open Project Program of Traction Power State Key Laboratory of Southwest Jiaotong University under Grant TPL2002, in part by the Open Project Program of State Key Laboratory of Mechanical Transmissions of Chongqing University under Grant SKLMT-KFKT-201803, and in part by the Liaoning BaiQianWan Talents Program.

ABSTRACT Broad learning system (BLS) is an effective and efficient incremental learning system without the deep architecture. It has strong feature extraction ability and high computational efficiency. However, it is greatly limited in the applicability of supervised learning. For the collected actual data, more data are unlabeled data and less data are labeled data. To overcome these problems, Fick's law assisted propagation (FLAP) is introduced into the BLS to propose a new semi-supervised classification algorithm, namely FLAP-BLS in this paper. In the FLAP-BLS, the FLAP has the labeled ability from the labeled examples to unlabeled examples, it is used to mark plenty of unlabeled samples by few labeled samples in order to obtain a large number of labeled samples and build the sample data matrix. Then an efficient incremental BLS without deep structure can effectively extract features from large-scale data, it is used to effectively classify the sample matrix. Finally, USPS, MNIST and NORB datasets are selected to validate the effectiveness of the FLAP-BLS. The experiment results show that the FLAP-BLS can effectively classify the few labeled samples and a large of unlabeled samples and obtain classification results with high accuracy, and it has faster classification speed, stronger generalization ability and better stability. The proposed method provides a new method for image classification.

INDEX TERMS Semi-supervised learning, image classification, broad learning system, Fick's law assisted propagation.

I. INTRODUCTION

Classification includes signal classification [1], [2], image classification [3]–[5], mail classification [6], [7] and so on [8]–[11]. The essence of classification is to determine the categories of data. Broad learning system (BLS) is an efficient incremental learning system proposed [12]. It is an effective algorithm for solving the classification problem. In recent years, many scholars have used BLS to solve various classification problems. Zhao *et al.* [13] combined BLS with PCA to realize bearing classification. Zheng *et al.* [14] combined VMD with BLS for fault classification. The BLS and fuzzy theory are combined to solve the problem of multi classification [15]. Jin and Chen [16] combined BLS with regularization for classifying imbalance data. Jin *et al.* [17] proposed GBLS for image classification. Wang *et al.* [18] applied BLS to the emotional classification. Although the BLS is widely applied in various fields, it is mainly used in supervised classification tasks. In some cases, such as text classification [19]–[21], image classification [22]–[27], the obtained plenty of labeled samples is time consuming and expensive, while a lot of unlabeled samples and few labeled samples are easy to be collected. Therefore, most of the existing BLS are often limited by the number of label

The associate editor coordinating the review of this manuscript and approving it for publication was Imran Sarwar Bajwa.

samples, which result in the unsatisfactory classification effect. In addition, other methods have been proposed in recent years [28]–[30].

Semi-supervised learning is a learning method based on combining supervised learning with unsupervised learning, which has capability for processing unlabeled samples and limited labeled samples simultaneously. In recent years, many scholars have proposed many graph-based semi-supervised learning methods. The adjacency structure of graph is constructed by KNN neighborhood, which further determines weight matrices by Gaussian kernel [31], [32], non-negative local linear reconstruction coefficients [33], and so on. However, the above semi supervised learning method has many shortcomings. Considering these problems, SSL methods based on sparse graph were successively proposed. The nonnegative low-rank and sparse graph proposed by Zhuang *et al.* [34] can capture both the global mixture of subspaces structure and the locally linear structure of data, hence it is both generative and discriminative. de Morsier *et al.* [35] presented a kernel low-rank and sparse graph, which was based on sample proximities in reproducing kernel Hilbert spaces and expressed sample relationships under sparse and low-rank constraints. However, data class structure is not considered in the above methods. Considering this, Gong *et al.* [36] proposed a Fick's law assisted propagation (FLAP) algorithm. It can reflect the relationship between samples and each category through the label information matrix. More importantly, the FLAP can accurately spread the information of the labeled samples to the unlabeled samples with linear speed. In addition, a lot of optimization algorithm are proposed [28], [37]–[40], which can be used to optimize the parameters of classification models.

In practical applications, data often consist of a few numbers of labeled data and a plenty of unlabeled data, which result in low classification accuracy using BLS. The FLAP has the labeled ability from the labeled examples to unlabeled examples, it can obtain pseudo labels of unlabeled samples by expanding few labeled samples to the number of labeled samples. The BLS has simple structure and can effectively extract features from large-scale data. In this paper, the FLAP and BLS are combined to propose a new semi-supervised classification (FLAP-BLS) method. The FLAP-BLS inherits the characteristics of BLS and FLAP to expand the application field of the BLS. In addition, the FLAP-BLS can build feature nodes and enhancement nodes to extract features from data. Above all, the FLAP-BLS can use plenty of unlabeled samples and a few labeled samples to achieve semi-supervised classification with faster calculation speed and stronger generalization ability. The validity of the FLAP-BLS is verified by three different complex experiment data.

II. BASIC METHODS

A. FLAP

FLAP is a label propagation method to simulate fluid diffusion. The propagation process of FLAP is constrained by Fick's first law of diffusion. Firstly, the label information spreads from labeled samples to unlabeled samples. Secondly, the amount of label information of unlabeled samples will be influenced by the labeled samples and other unlabeled samples. Next, the amount of label information of labeled samples will also be affected by other samples [41]. Finally, when the amount of label information of each sample reaches a balance, the propagation will end.

Denote the labeled samples in the training set as $\{X_l, Y_l\} = \{x_i, y_i\}_{i=1}^{l}$, and unlabeled data as $X_u = \{x_i\}_{i=1}^{u}$, where l and u are the number of labeled and unlabeled samples, respectively. $n = l + u$ is the number of all samples. $Y_l \in R^{l\times C}$ is the label of the labeled sample, C is the number of categories. At the beginning of information propagation, label information is transmitted from labeled samples to unlabeled samples.

$$f_j^{(t+1)} = f_j^{(t)} - \gamma \frac{f_j^{(t)} - f_i^{(t)}}{d_{ij}^2} \tag{1}$$

where $f_j^{(t+1)}$ is the amount of label information contained in sample x_j at time $t+1$. $f_j^{(t)}$ and $f_i^{(t)}$ represent the amount of information contained in samples x_j and x_i at time t, respectively. d_{ij} is the diffusion distance between samples x_i and x_j. If x_i is K-nearest neighbor of x_j, $d_{ij} = 1/\exp(-||x_i - x_j||^2/2\sigma^2)$, otherwise, $d_{ij} = \infty$. γ is the propagation coefficient, $i = 1, 2, \ldots, l, j = 1, 2, \ldots, u$.

In the process of label information propagation, the amount of label information of unlabeled samples will be affected by other samples. Equation (1) is transformed into the equation.

$$f_j^{(t+1)} = f_j^{(t)} - \sum_{k=1}^{n} \gamma \frac{f_j^{(t)} - f_k^{(t)}}{d_{kj}^2} \tag{2}$$

where n represents the number of all samples.

With the progress of information propagation, the interaction between samples and the influence of initial state is considered, the equation (2) is transformed into the following equation.

$$f_j^{(t+1)} = \alpha \left(f_j^{(t)} - \sum_{k=1}^{n} \gamma \frac{f_j^{(t)} - f_k^{(t)}}{d_{kj}^2} \right) + (1-\alpha) y_j \tag{3}$$

where $\alpha \in (0, 1)$ is compromise parameter, y_j represents the initial label information contained in the samples, $j = 1, 2, \ldots, n$.

The equation (3) is transformed into the following equation (4)and (5), as shown at the bottom of the next page. where $F^{(t+1)} = (f_1^{(t+1)}, f_2^{(t+1)}, \ldots, f_n^{t+1})^T$ and $F^{(t)} = (f_1^{(t)}, f_2^{(t)}, \ldots, f_n^{t})^T$ represent the information amount of all labels of all samples at time $t+1$ and time t.$Y = (y_1, y_2, \ldots, y_n)^T$ indicates the initial label information of all samples. $F^{(t)}$, $F^{(t+1)}$ and Y are of size $n \times C$, and C denotes the total number of categories. If x_j is labeled sample and the category is c', then $Y_{jc'} = 1$, the rest is 0. If x_j is unlabeled sample, $Y_{jc'} = 0$. P is a nonnegative random matrix,

the sum of elements in each row is 1, and γ is the propagation coefficient. $j = 1, 2, \ldots, n$, $c' = 1, 2, \ldots, C$.

When the label information concentration among samples meets the following conditions, the propagation will end.

$$||F^{(t+1)} - F^{(t)}||_F < \varepsilon \tag{6}$$

where $||\cdot||_F$ denotes Frobenius standard, ε is the number of errors.

The equation (4) and equation (6) are combined, the label information concentration of each sample is obtained.

$$F^* = (1-\alpha)(I - \alpha P)^{-1} Y \tag{7}$$

where $F^* \in R^{n \times C}$ represents the final label information of each sample.

Finally, the pseudo label of unlabeled samples is obtained according to the following equation.

$$\mathrm{Y}_j^u = argmax_{1 \le c' \le C} F_{jc'} \tag{8}$$

$$Y^U = (\mathrm{Y}_1^u, \mathrm{Y}_2^u, \ldots, \mathrm{Y}_u^u)^T \tag{9}$$

where Y_j^U is the pseudo label of the jth unlabeled sample, Y^U is the pseudo label matrix of all unlabeled samples, $j = 1, 2, \ldots, u$.

B. BLS

BLS is an effective and efficient incremental learning system without the deep architecture. Compared with deep networks, the structure of the BLS is simple, it consists of feature nodes, enhancement nodes and output coefficient matrix. It can effectively extract features from large-scale data by establishing feature nodes and enhancing nodes, so as to maintain the validity of the system for data.

The construction process of BLS is described as follows.

(1) For the input data X, the linear transformation function mapping is used to generate the ith set of mapping features Z_i.

$$Z_i = \phi_i(XW_{ei} + \beta_{ei}) \tag{10}$$

where w_{ei} and β_{ei} are random weight coefficients with appropriate dimensions and n is the number of groups of feature nodes, $i = 1, \ldots, n$. The number of feature nodes in each group is k, $\phi(\cdot)$ indicates linear transformation.

All feature nodes are combined and recorded them as Z^n.

$$Z^n \equiv [Z_1, \ldots, Z_n] \tag{11}$$

where $Z^n \in R^{(N \times nk)}$, N is the number of samples, nk is the number of all feature nodes.

(2) All feature node groups Z^n are generated according to the equation (11) are subjected to the nonlinear function transformation to generate enhancement node H_j.

$$H_j = \xi_j(Z^n W_{h_j} + \beta_{h_j}) \tag{12}$$

where W_{h_j} and β_{h_j} are random weight coefficients with appropriate dimensions, m is the total number of enhancement nodes, and $j = 1, \ldots, m$, $\xi(\cdot)$ is used for nonlinear function transformation.

(3) All enhancement nodes are defined as H^m.

$$H^m \equiv [H_1, \ldots, H_m] \tag{13}$$

where $H^m \in R^{(N \times m)}$, m is the total number of enhancement nodes.

The enhancement nodes and feature nodes are combined by equations (12) and (13) to obtain A.

$$A = [Z^n | H^m] \tag{14}$$

where $A \in R^{N \times (nk+m)}$, N is the number of samples, $nk + m$ is the number of all nodes.

(4) The BLS aims to solve the output coefficient by minimizing the sum of the square loss of prediction error.

$$\underset{\beta}{\operatorname{argmin}}\, \theta \|\beta\|^2 + \|A\beta - Y\|^2 \tag{15}$$

where Y is the label corresponding to data X, β is the output coefficient. θ is the regularization parameter, which is used to balance the error and model complexity.

The equation (15) is set as zero relative to the gradient.

$$\beta = (A^T A + \theta I)^{-1} A^T Y \tag{16}$$

where I is the unit matrix. If $\theta = 0$, the equation (15) degenerates into the least square problem. On the other hand, if $\theta \to \infty$, the solution is heavily constrained and tends to 0. Thus, we set $\theta \to 0$ here.

The structure of BLS is shown in Figure 1.

$$F^{(t+1)} = \alpha P F^{(t)} + (1-\alpha) Y \tag{4}$$

$$P = \begin{pmatrix} 1-\gamma \sum_{k=1,k\neq 1}^{n} d_{1k}^{-2} & \gamma d_{12}^{-2} & \cdots & \gamma d_{1n}^{-2} \\ \gamma d_{21}^{-2} & 1-\gamma \sum_{k=1,k\neq 2}^{n} d_{2k}^{-2} & \cdots & \gamma d_{2n}^{-2} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma d_{n1}^{-2} & \gamma d_{n2}^{-2} & \cdots & 1-\gamma \sum_{k=1,k\neq 1}^{n} d_{nk}^{-2} \end{pmatrix} \tag{5}$$

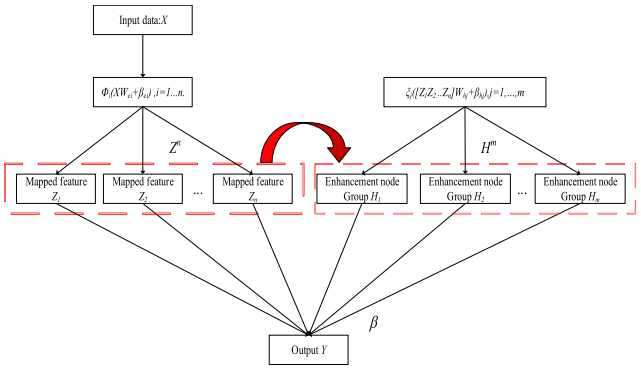


FIGURE 1. The structure of the BLS.

III. A SEMI-SUPERVISED CLASSIFICATION METHOD

In practical engineering, data often consist of a few numbers of labeled data and plenty of unlabeled data. The BLS cannot better classify the unlabeled data. The FLAP can spread the label information of labeled samples to unlabeled samples, and obtain pseudo labels of unlabeled samples, and expand the number of labeled samples. To improve the classification performance of BLS, a new semi-supervised classification method based on FLAP and BLS, namely FLAP-BLS is proposed in this paper. In the FLAP-BLS, the FLAP use a few number of labeled samples to mark the unlabeled samples to expand the number of labeled samples and build the data sample matrix. Secondly, the labeled samples and unlabeled samples are used to build feature nodes and enhancement nodes. Thirdly, the objective function is constructed by feature nodes, enhancement nodes, output coefficient, labels of labeled samples and pseudo labels of unlabeled samples. Finally, the objective function is solved by ridge regression, and the output coefficient is obtained. The FLAP-BLS can obtain pseudo labels of unlabeled samples and expand the number of labeled samples. The feature nodes and enhancement nodes can realize feature extraction of samples. The output coefficient can be obtained by ridge regression, which ensures the learning efficiency.

In the process of training, the input data X consists of labeled data sets $\{X^l, Y^l\} = \{x^i, y^i\}_{i=1}^{l}$ and unlabeled data sets $X^u = \{x^i\}_{i=1}^{u}$, where l and u are the number of labeled data and unlabeled data, respectively. Y^l is the label corresponding to the labeled sample. Y^U is the pseudo label corresponding to the unlabeled sample, which is obtained by the equation (9).

(1) For the input data X, the linear transformation function mapping is used to generate the ith set of mapping features Z_i.

$$Z_i = \phi_i(XW_{ei} + \beta_{ei}) \tag{17}$$

where w_{ei} and β_{ei} are random weight coefficients with appropriate dimensions and n is the number of groups of feature nodes, $i = 1, \ldots, n$. The number of feature nodes in each group is k, $\phi(\cdot)$ indicates linear transformation.

All feature nodes are combined and recorded them as Z^n.

$$Z^n \equiv [Z_1, \ldots, Z_n] \tag{18}$$

where $Z^n \in R^{(l+u)\times nk}$, $l+u$ is the number of samples, nk is the number of all feature nodes.

(2) All feature node groups Z^n are generated according to the equation (18) are subjected to the nonlinear function transformation to generate enhancement node H_j.

$$H_j = \xi_j(Z^n W_{h_j} + \beta_{h_j}) \tag{19}$$

where W_{h_j} and β_{h_j} are random weight coefficients with appropriate dimensions, m is the total number of enhancement nodes, and $j = 1, \ldots, m$, $\xi(\cdot)$ is used for nonlinear function transformation.

(3) All enhancement nodes are defined as H^m.

$$H^m \equiv [H_1, \ldots, H_m] \tag{20}$$

where $H^m \in R^{(l+u)\times m}$, m is the total number of enhancement nodes.

The enhancement nodes and feature nodes are combined by equations (18) and (20) to obtain A.

$$A = [Z^n | H^m] \tag{21}$$

where $A \in R^{(l+u)\times(nk+m)}$, $(l+u)$ is the number of samples, $nk+m$ is the number of all nodes.

(4) The BLS aims to solve the output coefficient by minimizing the sum of the square loss of prediction error.

$$\underset{\beta}{\operatorname{argmin}}\, \theta \|\beta\|^2 + \left\| A\beta - [Y^l | Y^U] \right\|^2 \tag{22}$$

where β is the output coefficient. θ is the regularization parameter, which is used to balance the error and model complexity.

The equation (22) is set as zero relative to the gradient.

$$\beta = (A^T A + \theta I)^{-1} A^T [Y^l | Y^U] \tag{23}$$

where I is the unit matrix. If $\theta = 0$, the equation (22) degenerates into the least square problem. On the other hand, if $\theta \to \infty$, the solution is heavily constrained and tends to 0. Thus, we set $\theta \to 0$ here.

The model of the FLAP-BLS is shown in Figure 2.

The main flow of the FLAP-BLS is shown in Algorithm 1.

IV. VALIDATION AND ANALYSIS

A. EXPERIMENT DATA AND ENVIRONMENT

To verify the performance of the FLAP-BLS, the USPS, MNIST and NORB data sets are selected in here. The FLAP-BLS is compared with the FLAP-SVM, FLAP-KELM, FLAP-HELM, LPDGL-BLS and TLLT-BLS. FlAP-SVM based on FLAP and support vector machine (SVM) [34], FLAP-KELM based on FLAP and kernel extreme learning machine (KELM) [42]–[44], FLAP-HELM based on FLAP and hierarchical extreme learning machine (HELM), LPDGL-BLS based on label propagation of deformed graph laplacian (LPDGL) and BLS, TLLT-BLS

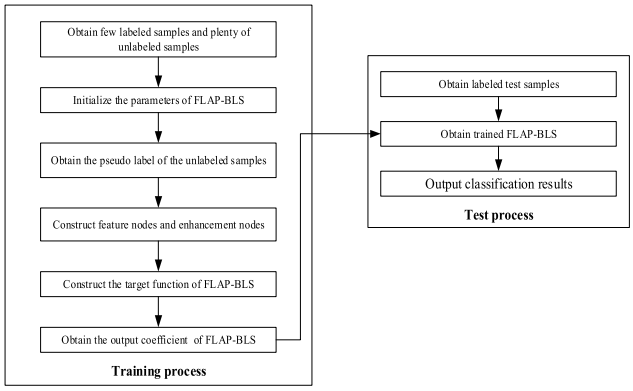


FIGURE 2. The model of the FLAP-BLS.

Algorithm 1 FLAP-BLS Algorithm

Input

Labeled samples, $\{X^l, Y^l\} = \{x^i, y^i\}_{i=1}^{l}$, unlabeled samples, $X^u = \{x^i\}_{i=1}^{u}$

Output

output coefficient β

Step 1. Select the nearest neighbor number K, propagation coefficient γ, compromise parameter α, error number ε, node number k, characteristic node group number n, enhancement node number m, regularization parameter θ.

Step 2. Obtaining pseudo labels Y^U of unlabeled samples through labeled samples $\{X^l, Y^l\}$ and unlabeled samples X^u.

Step 3. Construct the feature node layer and the enhancement node layer by using the weights and deviations of the random input and calculate the output matrix $A \in R^{(l+u)\times(nk+m)}$.

Step 4. The objective function is constructed by output matrix A, output coefficient β, labels Y^l of labeled samples, and pseudo labels Y^u of unlabeled samples.

Step 5. Calculate the output coefficient β by ridge regression.

based on Teaching-to-Learn and Learning-to-Teach for label Propagation (TLLT) and BLS.

The regularization parameters of SVM, KELM, HELM and BLS are 10^2, 10^2, 10^{-8} and 10^{-10}, respectively. The kernel parameters of SVM and KELM are 10^2. The structure of FLAP-BLS and FLAP-HELM is set as 10-10-1000. The nearest neighbor number, propagation coefficient, compromise parameter and error number of FLAP are 10, 0.1, 0.9 and 10^{-5} respectively. The nearest neighbors and equilibrium parameters of LPDGL are 10 and 1. The neighbor number, learning rate, evaluation parameters and reliability trade-off parameters of TLLT are 10, 2, 1, and 0.5 respectively. In addition, all experiments are tested by MATLAB 2018b on Intel-i5 2.4GHz CPU and 8GB memory.

TABLE 1. Comparison results of classification (100-298).

Methods	Test accuracy (%)	Classification time (s)	STD
SVM	65.86	0.09	**2.15**
KELM	70.65	**0.07**	4.51
HELM	68.65	0.20	2.25
BLS	**73.60**	0.24	3.94

TABLE 2. Comparison results of classification (100-8500-298).

Methods	Test accuracy (%)	Classification time (s)	STD
FLAP-SVM	87.06	98.43	0.35
FLAP-KELM	81.27	94.51	0.83
FLAP-HELM	82.13	85.51	0.86
FLAP-BLS	**88.95**	**85.71**	**0.33**

TABLE 3. Comparison results of classification (100-8500-298).

Methods	Test accuracy (%)	Classification time (s)	STD
LPDGL-BLS	84.28	505.90	0.67
TLLT-BLS	87.41	2288.21	0.56
FLAP-BLS	**88.95**	**85.71**	**0.33**

In here, 9298 data samples are selected from three data sets. 9298 data samples were divided into 500 labeled training samples, 8500 unlabeled training samples and 298 labeled test samples. Each method runs 10 times independently. The average classification time, average test accuracy and test accuracy standard deviation (STD) are selected as comparison indexes.

B. USPS DATA

The dataset of USPS is used to test the efficiency of the FLAP-BLS in here. The digital set contains 9298 images and 10 categories. The resolution of each image is 16×16 grayscale pixels. The experiment results are shown in Tab.1, Tab.2 and Tab.3.

From Tab.1 and Tab. 2, the test accuracy and STD value of BLS are 73.60% and 3.94s, respectively. The test accuracy and STD of FLAP-BLS are 88.95% and 0.33s, respectively. The experiment results show that the FLAP-BLS can effectively use unlabeled samples to build a semi-supervised classification model with strong stability and high classification accuracy. Compared with FLAP-HELM, FLAP-SVM and FLAP-KELM, the FLAP-BLS can obtain the best classification accuracy of 88.95%. The classification time of the FLAP-HELM is 85.51s, which is least running time. The classification time of the FLAP-BLS is 85.71s, which is also much faster than that of the FLAP-SVM and FLAP-KELM. And the classification time of the FLAP-HELM and FLAP-BLS are minimal and almost equal.

From Tab.3, for the classification efficiency, the TLLT-BLS is lowest and the FLAP-BLS is highest. For the

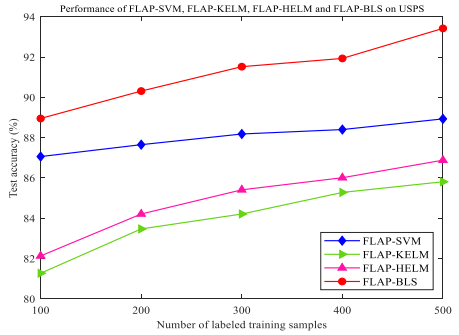


FIGURE 3. The test accuracy with increasing of labeled training samples.

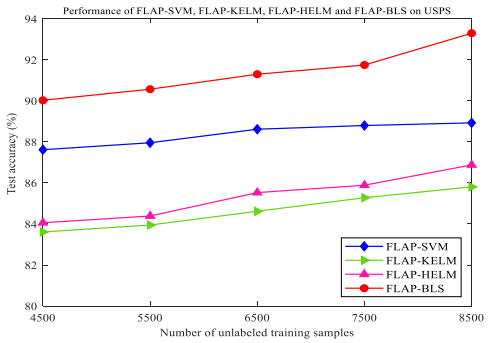


FIGURE 4. The test accuracy with increasing of unlabeled training samples.

FLAP-BLS, the test accuracy is 88.95%, which is better in these methods. The experiment results show that the FLAP-BLS has better performance than the LPDGL-BLS and TLLT-BLS.

To analyze the influence of the number of labeled training samples on the USPS, the number of unlabeled samples is 8500, the number of labeled test samples are 298. Then the number of labeled training samples increases gradually with 100 labeled samples, up to 500. The test accuracy is shown in Figure 3.

From the Figure 3, with increasing of labeled training samples, the test accuracies of FLAP-SVM, FLAP-KELM, FLAP-HELM and FLAP-BLS increase gradually. Under the same number of labeled training samples, the FLAP-BLS has better test accuracy than FLAP-SVM, FLAP-KELM and FLAP-HELM.

To test the influence of the number of unlabeled training samples on the USPS, the number of labeled training samples is 500, the number of labeled test samples are 298, the initial number of unlabeled training samples is 4500. Then the number of unlabeled training samples increases gradually with 1000 unlabeled training samples, up to 8500. The test accuracy is shown in Figure 4.

From the Figure 4, with increasing of unlabeled training samples, the test accuracies of FLAP-SVM, FLAP-KELM, FLAP-HELM and FLAP-BLS increase gradually. Under the same number of unlabeled training samples, the FLAP-BLS can obtain best test accuracy.

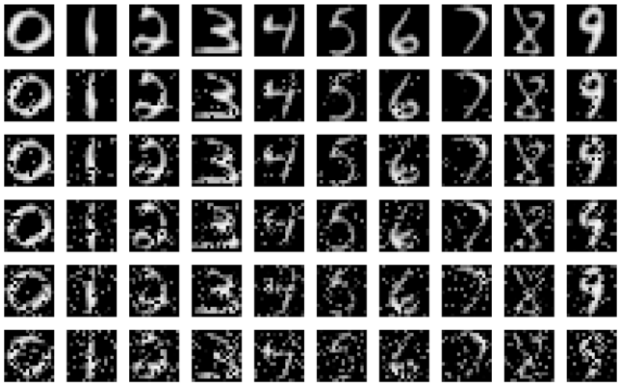


FIGURE 5. MNIST data samples.

TABLE 4. Comparison results of classification (100-298).

Methods	Test accuracy (%)	Classification time (s)	STD
SVM	73.31	0.29	**2.58**
KELM	73.92	0.17	3.14
HELM	71.93	**0.13**	4.18
BLS	**75.47**	0.30	4.17

TABLE 5. Comparison results of classification (100-8500-298).

Methods	Test accuracy (%)	Classification time (s)	STD
FLAP-SVM	82.18	159.43	0.34
FLAP-KELM	**85.95**	**85.51**	**0.83**
FLAP-HELM	78.71	86.51	0.86
FLAP-BLS	85.84	87.31	0.33

In summary, it is not difficult to see from the experiment results that whether the number of labeled training samples or the number of unlabeled training samples are increased gradually, the test accuracies of the FLAP-SVM, FLAP-KELM, FLAP-HELM and FLAP-BLS are improved gradually. At the same time, under the same number of labeled training samples or unlabeled training samples, the FLAP-BLS can obtain higher semi-supervised classification accuracy and efficiency.

C. MNIST DATA

To further testify the effectiveness of the FLAP-BLS, MNIST data is selected in here. MNIST data are handwritten digital images with consisting of 70,000 handwritten digits. Each number is represented by images with 28 × 28 grayscale pixels. MNIST data samples are shown in Figure 5.

The experiment results are shown in Tab.4, Tab.5 and Tab.6.

From Tab.4 and Tab.5, the test accuracy of BLS is 75.47%, and the test accuracy of FLAP-BLS is 85.84%. The STD of FLAP-BLS is 0.33. the test accuracy of KELM is 73.92%, and the test accuracy of FLAP-KELM is 85.95%. The STD

TABLE 6. Comparison results of classification (100-8500-298).

Methods	Test accuracy (%)	Classification time (s)	STD
LPDGL-BLS	83.94	462.56	0.51
TLLT-BLS	80.27	4202.69	0.46
FLAP-BLS	**85.84**	**87.31**	**0.33**

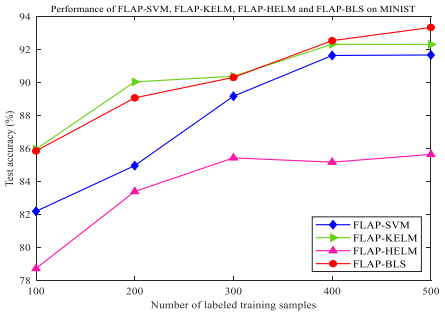


FIGURE 6. The test accuracy with increasing of labeled training samples.

of FLAP-KELM is 0.83. The experiment results show that the FLAP can effectively expand the number of label samples and improve the accuracy and stability of the classification method. The test accuracy of FLAP-BLS is better than that of FLAP-SVM and FLAP-HELM, and is slightly lower than that of FLAP-KELM. The STD of the FLAP-BLS are better than those of the FLAP-SVM, FLAP-KELM, and FLAP-HELM.

From Tab.6, the test accuracy, average classification time and STD of the FLAP-BLS are 85.84%, 87.31s and 0.33, respectively, which are better than those of the LPDGL-BLS and TLLT-BLS. Therefore, the experiment results show that the FLAP-BLS has better classification effect on MNIST data set.

To analyze the influence of the number of labeled training samples on the MNIST for the FLAP-BLS, the number of unlabeled samples is 8500, the number of initial labeled training samples are 100, and the number of labeled test samples are 298. Then the number of labeled training samples increases gradually with 100 labeled training samples, up to 500. The test result is shown in Figure 6.

With increasing of labeled training samples, the test accuracies of the FLAP-SVM and FLAP-HELM are improved gradually, while the FLAP-BLS increase firstly and then be stable. When the number of labeled training samples increases from 100 to 300, the test accuracy of the FLAP-BLS is slightly lower than that of the FLAP-KELM. When the labeled training samples are increased from 300 to 500, the test accuracy of the FLAP-BLS is higher than that of the FLAP-KELM. In general, the FLAP-BLS can better realize semi-supervised classification for the MNIST data.

To test the influence of the number of unlabeled training samples on the MNIST for the FLAP-BLS, the number of labeled training samples is 500, the number of labeled test samples are 298, the number of initial unlabeled training samples are 4500. Then the number of unlabeled training

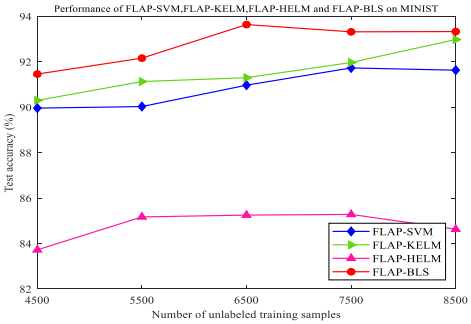


FIGURE 7. The test accuracy with increasing of unlabeled training samples.

TABLE 7. Comparison results of classification (100-298).

Methods	Test accuracy (%)	Classification time (s)	STD
SVM	58.74	0.69	6.02
KELM	60.86	**0.08**	**4.42**
HELM	60.73	**0.08**	6.43
BLS	**64.43**	0.33	4.71

samples increases gradually with 1000 unlabeled training samples, up to 8500. The test accuracy is shown in Figure 7.

With increasing of unlabeled training samples, the test accuracy of the FLAP-KELM is improved gradually, while FLAP-SVM, FLAP-HELM, and FLAP-BLS increase firstly and then decrease. But the final test accuracies of the FLAP-SVM, FLAP-HELM and FLAP-BLS still are better than their initial test accuracies. Under the same number of labeled training samples or unlabeled training samples, the FLAP-BLS can meet the requirements of semi-supervised classification and obtain best test accuracy.

In summary, it is not difficult to see from the experiment results that the number of labeled training samples is increased for MNIST data, test accuracies of the FLAP-BLS are improved. With increasing of unlabeled training samples, the test accuracy of the FLAP-BLS increase firstly and then decrease. At the same time, under the same number of labeled training samples or unlabeled training samples, the FLAP-BLS can better obtain the classification results of MNIST data.

D. NORB DATA

Compared to the MNIST data, NORB data is a more complex data. It includes 48,600 images with $2 \times 32 \times 32$ pixels. The NORB data contains images with 5 different categories, which are 50 different 3D toy objects. 5 different categories are animals, humans, aircrafts, trucks and cars. The sampled objects under various lighting conditions are shown in Figure 8.

The experiment results are shown in Tab.7, Tab.8, and Tab.9.

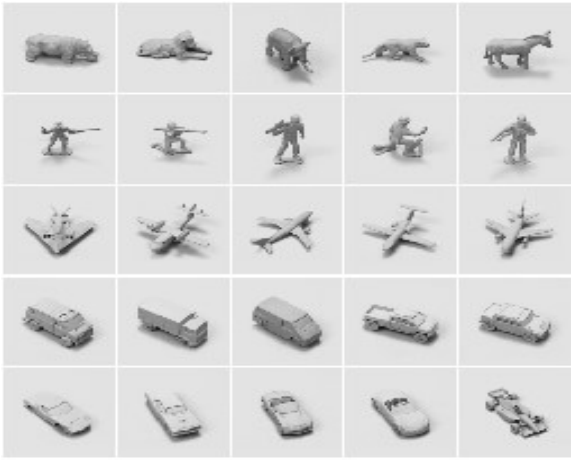


FIGURE 8. The sampled objects of NORB data.

TABLE 8. Comparison results of classification (100-8500-298).

Methods	Test accuracy (%)	Classification time (s)	STD
FLAP-SVM	68.32	242.43	1.51
FLAP-KELM	71.90	95.51	1.68
FLAP-HELM	73.37	86.51	0.66
FLAP-BLS	**74.35**	**89.47**	**0.51**

TABLE 9. Comparison results of classification (100-298).

Methods	Test accuracy (%)	Classification time (s)	STD
LPDGL-BLS	56.61	522.61	2.21
TLLT-BLS	70.24	12438.58	1.72
FLAP-BLS	**74.35**	**89.47**	**0.51**

From Tab.7 and Tab.8, the test accuracy and STD of the BLS are 64.43% and 4.71, respectively. The test accuracy and STD of the FLAP-BLS are 74.35% and 0.51, respectively. For the FLAP-BLS, the test accuracy is the highest in these classification methods. The experiment results show that the FLAP-BLS can effectively expand the number of training samples, improve classification accuracy and enhance stability. It can effectively complete semi-supervised classification task for NORB data with better test accuracy and faster classification time.

From Tab.9, the test accuracy of LPDGL-BLS is 56.61%, which is the worst of the three methods. The test accuracy, classification time, and STD of FLAP-BLS are 74.35%, 89.47s and 0.51, respectively, which are superior to the LPDGL-BLS, and TLLT-BLS. The test accuracy of the FLAP-BLS is more stable and the modeling time of the FLAP-BLS is shorter than other methods. Therefore, the FLAP-BLS can effectively obtain classification results for the NORB data.

To validate the influence of the number of labeled training samples on the NORB data using the FLAP-BLS, the number

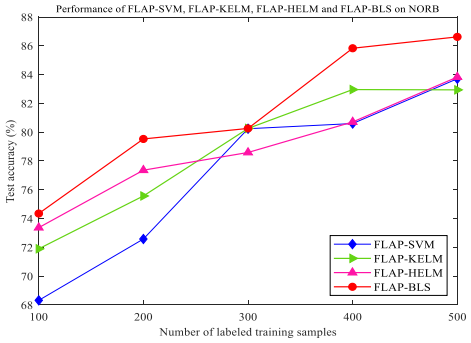


FIGURE 9. The test accuracy with increasing of labeled training samples.

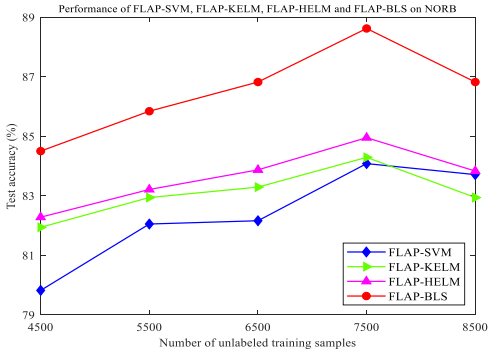


FIGURE 10. The test accuracy with increasing of unlabeled training samples.

of unlabeled samples is 8500, the number of initial labeled training samples are 100, and the number of labeled test samples are 298. Then the number of labeled training samples increases gradually with 100 labeled training samples, up to 500. The test accuracy is shown in Figure 9.

From Figure 9, with increasing of labeled training samples, the test accuracies of FLAP-SVM, FLAP-KELM and FLAP-HELM are improved gradually. Under the same number of labeled training samples, the test accuracy of the FLAP-BLS is best. Therefore, the FLAP-BLS is superior to the FLAP-SVM, FLAP-KELM, and FLAP-HELM.

To test the influence of the number of unlabeled training samples on the NORB using the FLAP-BLS, the number of labeled training samples is 500, the number of labeled test samples are 298, the number of initial unlabeled training samples are 4500. The number of unlabeled training samples increases gradually with 1000 unlabeled training samples, up to 8500. The test accuracy is shown in Figure 10.

From Figure 10, with increasing of unlabeled training samples, the test accuracies of the FLAP-SVM, FLAP-KELM, FLAP-HELM and FLAP-BLS increase firstly and then decrease. But the final test accuracies of the FLAP-SVM, FLAP-KELM, FLAP-HELM and FLAP-BLS still are better than their initial test accuracies. In practical applications, the FLAP-BLS with the number of unlabeled samples can be flexibly choose according to different practical requirements. At same time, under the same number of unlabeled training samples, the FLAP-BLS can obtain best test accuracy.

In summary, it is not difficult to see from the experiment results that the test accuracies of the FLAP-SVM, FLAP-KELM, FLAP-HELM and FLAP-BLS increase firstly and then decrease with increasing of the number of labeled training samples. At the same time, under the same number of unlabeled training samples or unlabeled training samples, the FLAP-BLS can meet the requirements of semi-supervised classification with best test accuracy. The experimental results show that the FLAP-BLS method has good classification performance for NORB data.

E. EXPERIMENT RESULT ANALYSIS

From the experiment results of USPS data, with increasing of unlabeled training samples, the test accuracy of the FLAP-BLS improve gradually. As can be seen from the experiment classification results of MNIST and NORB data, with increasing of unlabeled training samples, the test accuracy of FLAP-BLS increase first and then decrease. This is show that the propagation efficiency of FLAP is relate to the complexity of data. And the complex of samples is larger, the propagation efficiency is lower.

The FLAP-BLS can obtain better classification results under semi-supervised environments with different complex data. The exepriment results show that the FLAP-BLS can more effectively extract data features from different data in order to ensure the test accuracy and reduce the training time of FLAP-BLS. At the same time, it also shows that the FLAP-BLS has high stability and strong adaptability. It is more adaptable to semi-supervised environments of different data than other semi supervised classification algorithms.

V. CONCLUSIONS AND PROSPECTS

In this paper, a new semi-supervised classification algorithm based on FLAP and BLS, namely FLAP-BLS is proposed. The FLAP can transmit label information of labeled samples to unlabeled samples. The BLS has strong learning ability, simple structure and fast calculation speed, which can efficiently realize the classification. USPS, MNIST and NORB datasets are used to verify the validity of the FLAP-BLS method. Compared with BLS, the FLAP-BLS can classify few labeled samples and plenty of unlabeled samples and obtain ideal classification results. Compared with TLLT-BLS and LPDGL-BLS, the FLAP-BLS can realize the semi-supervised classification with the high accuracy and less computing time. Compared with FLAP-SVM, FLAP-KELM, and FLAP-HELM, the FLAP-BLS has better classification efficiency and higher classification accuracy. More importantly, the FLAP-BLS can obtain better classification effect on different complex datasets, which shows that the FLAP-BLS method has strong generalization ability. Therefore, the FLAP-BLS has higher classification accuracy, faster classification speed, stronger generalization ability and better stability.

However, the FLAP-BLS still has own limitations. When it faced with complex data set, such as NORB data sets, the test efficiency and test accuracy are not ideal enough. To solve the problem, a new semi-supervised classification method will be deeply studied for complex data in future. Moreover, the values of the hyperparameters (including the nearest neighbor number, the propagation coefficient, compromise parameter and error number) are selected according to the empirical values. How to efficiently select hyperparameters of the FLAP-BLS is also our next research work.

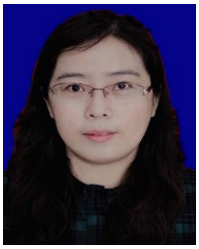
REFERENCES

[1] S. Banerjee and M. Mitra, "Application of cross wavelet transform for ECG pattern analysis and classification," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 2, pp. 326–333, Feb. 2014.
[2] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network," *IEEE Trans. Instrum. Meas.*, early access, Mar. 25, 2020, doi: 10.1109/TIM.2020.2983233.
[3] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1259–1270, Mar. 2018.
[4] H. Zhao, J. Zheng, W. Deng, and Y. Song, "Semi-supervised broad learning system based on manifold regularization and broad network," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 3, pp. 983–994, Mar. 2020.
[5] G. Tsagkatakis, M. Bloemen, B. Geelen, B. Geelen, M. Jayapala, and P. Tsakalides, "Graph and rank regularized matrix recovery for snapshot spectral image demosaicing," *IEEE Trans. Comput. Imag.*, vol. 5, no. 2, pp. 301–316, Jun. 2019.
[6] E. H. Alaa, "Filtering spam e-mail from mixed arabic and English messages: A comparison of machine learning techniques," *Int. Arab. J. Inf. Techn.*, vol. 6, no. 1, pp. 52–59, 2009.
[7] X.-Y. Zhang, C. Li, H. Shi, X. Zhu, P. Li, and J. Dong, "AdapNet: Adaptability decomposing encoder-decoder network for weakly supervised action recognition and localization," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 23, 2020, doi: 10.1109/TNNLS.2019.2962815.
[8] T. Li, J. Shi, X. Li, J. Wu, and F. Pan, "Image encryption based on pixel-level diffusion with dynamic filtering and DNA-level permutation with 3D Latin cubes," *Entropy*, vol. 21, no. 3, p. 319, Mar. 2019.
[9] H. Chen, S. Jiao, A. A. Heidari, M. Wang, X. Chen, and X. Zhao, "An opposition-based sine cosine approach with local search for parameter estimation of photovoltaic models," *Energy Convers. Manage.*, vol. 195, pp. 927–942, Sep. 2019.
[10] H. Zhao, H. Liu, J. Xu, and W. Deng, "Performance prediction using high-order differential mathematical morphology gradient spectrum entropy and extreme learning machine," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4165–4172, Jul. 2020.
[11] W. Deng, J. J. Xu, and H. M. Zhao, "An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem," *IEEE Access*, vol. 7, pp. 20281–20292, 2019.
[12] C. L. P. Chen and Z. Liu, "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.
[13] H. Zhao, J. Zheng, J. Xu, and W. Deng, "Fault diagnosis method based on principal component analysis and broad learning system," *IEEE Access*, vol. 7, pp. 99263–99272, 2019.
[14] J. Zheng, Y. Yuan, L. Zou, W. Deng, C. Guo, and H. Zhao, "Study on a novel fault diagnosis method based on VMD and BLM," *Symmetry*, vol. 11, no. 6, p. 747, Jun. 2019.
[15] S. Feng and C. L. P. Chen, "Fuzzy broad learning system: A novel neuro-fuzzy model for regression and classification," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 414–424, Feb. 2020, doi: 10.1109/TCYB.2018.2857815.
[16] J.-W. Jin and C. L. P. Chen, "Regularized robust broad learning system for uncertain data modeling," *Neurocomputing*, vol. 322, pp. 58–69, Dec. 2018.
[17] J. Jin, Z. Liu, and C. L. P. Chen, "Discriminative graph regularized broad learning system for image recognition," *Sci. China Inf. Sci.*, vol. 61, no. 11, Nov. 2018, Art. no. 112209.
[18] H. X. Wang, T. Zhang, and M. X. Xu, "EEG emotion recognition using dynamical graph convolutional neural networks and broad learning system," in *Proc. BIBM*, Dec. 2918, pp. 1240–1244.
[19] J. Chen, H. Huang, S. Tian, and Y. Qu, "Feature selection for text classification with Naïve Bayes," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5432–5435, Apr. 2009.

[20] Y. Liu, Y. Mu, K. Chen, Y. Li, and J. Guo, "Daily activity feature selection in smart homes based on pearson correlation coefficient," *Neural Process. Lett.*, vol. 51, no. 2, pp. 1771–1787, Apr. 2020, doi: 10.1007/s11063-019-10185-8.
[21] 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, Nov. 2015.
[22] Y. Kong, X. Wang, Y. Cheng, and C. Chen, "Hyperspectral imagery classification based on semi-supervised broad learning system," *Remote Sens.*, vol. 10, no. 5, p. 685, Apr. 2018.
[23] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
[24] Y. Xue, B. Xue, and M. Zhang, "Self-adaptive particle swarm optimization for large-scale feature selection in classification," *ACM Trans. Knowl. Discov. Data*, vol. 13, no. 5, p. 50, 2019.
[25] H. Zhao, D. Li, W. Deng, and X. Yang, "Research on vibration suppression method of alternating current motor based on fractional order control strategy," *Proc. Inst. Mech. Eng. E, J. Process Mech. Eng.*, vol. 231, no. 4, pp. 786–799, Aug. 2017.
[26] X. Li, Z. Xie, J. Wu, and T. Li, "Image encryption based on dynamic filtering and bit cuboid operations," *Complexity*, vol. 2019, Feb. 2019, Art. no. 7485621.
[27] Y. Liu, X. Yi, R. Chen, Z. Zhai, and J. Gu, "Feature extraction based on information gain and sequential pattern for english question classification," *IET Softw.*, vol. 12, no. 6, pp. 520–526, Dec. 2018.
[28] Y. Xu, H. Chen, J. Luo, Q. Zhang, S. Jiao, and X. Zhang, "Enhanced moth-flame optimizer with mutation strategy for global optimization," *Inf. Sci.*, vol. 492, pp. 181–203, Aug. 2019.
[29] R. Chen, S.-K. Guo, X.-Z. Wang, and T.-L. Zhang, "Fusion of multi-RSMOTE with fuzzy integral to classify bug reports with an imbalanced distribution," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 12, pp. 2406–2420, Dec. 2019.
[30] W. Deng, W. Li, and X.-H. Yang, "A novel hybrid optimization algorithm of computational intelligence techniques for highway passenger volume prediction," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4198–4205, Apr. 2011.
[31] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
[32] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2003, pp. 321–328.
[33] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 55–67, Jan. 2008.
[34] L. Zhuang, S. Gao, J. Tang, J. Wang, Z. Lin, Y. Ma, and N. Yu, "Constructing a nonnegative low-rank and sparse graph with data-adaptive features," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3717–3728, Nov. 2015.
[35] F. de Morsier, M. Borgeaud, V. Gass, J.-P. Thiran, and D. Tuia, "Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3410–3420, Jun. 2016.
[36] C. Gong, D. Tao, K. Fu, and J. Yang, "Fick's law assisted propagation for semisupervised learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 2148–2162, Dec. 2015.
[37] Y. Liu, X. Wang, Z. Zhai, R. Chen, B. Zhang, and Y. Jiang, "Timely daily activity recognition from headmost sensor events," *ISA Trans.*, vol. 94, pp. 379–390, Nov. 2019.
[38] Y. Xu, H. Chen, A. A. Heidari, J. Luo, Q. Zhang, X. Zhao, and C. Li, "An efficient chaotic mutative moth-flame-inspired optimizer for global optimization tasks," *Expert Syst. Appl.*, vol. 129, pp. 135–155, Sep. 2019.
[39] W. Deng, J. Xu, Y. Song, and H. Zhao, "An effective improved co-evolution ant colony optimization algorithm with multi-strategies and its application," *Int. J. Bio-Inspired Comput.*, pp. 1–10, 2020.
[40] H. Fu, M. Wang, P. Li, S. Jiang, W. Hu, X. Guo, and M. Cao, "Tracing knowledge development trajectories of the Internet of Things domain: A main path analysis," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6531–6540, Dec. 2019.
[41] Y. Wu, X. Yang, A. Plaza, F. Qiao, L. Gao, B. Zhang, and Y. Cui, "Approximate computing of remotely sensed data: SVM hyperspectral image classification as a case study," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5806–5818, Dec. 2016.
[42] M. Z. Alom, P. Sidike, T. M. Taha, and V. K. Asari, "State preserving extreme learning machine: A monotonically increasing learning approach," *Neural Process. Lett.*, vol. 45, no. 2, pp. 703–725, Apr. 2017.
[43] T. Li, Z. Qian, and T. He, "Short-term load forecasting with improved CEEMDAN and GWO-based multiple kernel ELM," *Complexity*, vol. 2020, Feb. 2020, Art. no. 1209547.
[44] J. Tang, C. Deng, and G.-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 809–821, Apr. 2016.

JIANJIE ZHENG (Member, IEEE) received the B.S. degree in soft engineering from the Dalian Institute of Science and Technology, Dalian, China, in 2018. He is currently pursuing the master's degree with Dalian Jiaotong University, Dalian. His research interest includes artificial intelligence.

YU YUAN (Member, IEEE) received the B.S. and M.S. degrees in internal combustion engine engineering and the Ph.D. degree in mechatronic engineering from the Dalian University of Technology, Dalian, China, in 2000, 2004, and 2007, respectively. Since 2011, he has been an Associate Professor with the Communication Equipment and Control Engineering, Dalian Jiaotong University, Dalian. His research interests include artificial intelligence, signal processing, and fault diagnosis.



HUIMIN ZHAO received the B.S. degree in electrical technology, the M.S. degree in traffic information engineering and control, and the Ph.D. degree in mechanical engineering from Dalian Jiaotong University, Dalian, China, in 2000, 2004, and 2013, respectively. Since 2017, she has been a Professor. Her research interests include artificial intelligence, signal processing, and fault diagnosis.

WU DENG (Member, IEEE) received the B.S. degree in electrical technology and the M.S. degree in computer application technology from Dalian Jiaotong University, Dalian, China, in 2001 and 2006, respectively, and the Ph.D. degree in computer application technology from Dalian Maritime University, Dalian, in 2012. Since 2019, he has been a Professor with the College of Electronic Information and Automation, Civil Aviation University of China, Tianjin, China. His research interests include artificial intelligence, optimization method, and fault diagnosis.