IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 30, 2018, accepted January 28, 2019, date of publication February 26, 2019, date of current version March 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901860

Semi-Supervised Deep Fuzzy C-Mean Clustering for Imbalanced Multi-Class Classification

ALI ARSHAD[1,2], SAMAN RIAZ[1,2], AND LICHENG JIAO[3], (Fellow, IEEE)

[1]School of Computer Science and Technology, Xidian University, Xi'an 710071, China
[2]School of International Education, Xidian University, Xi'an 710071, China
[3]International Joint Collaboration Laboratory of Intelligent Perception and Computation, International Research Center of Intelligent Perception and Computation, Key Laboratory of Intelligent Perception and Image Understanding, School of Artificial Intelligence, Ministry of Education of China, Xidian University, Xi'an 710071, China

Corresponding author: Ali Arshad (alli.arshad@gmail.com)

This work was supported in part by the National Basic Research Program (973 Program) of China under Grant 2013CB329402, in part by the National Natural Science Foundation of China under Grant 61573267, Grant 61473215, Grant 61571342, Grant 61572383, Grant 61501353, Grant 61502369, Grant 61271302, Grant 61272282, and Grant 61202176, in part by the Fund for Foreign Scholars in University Research and Teaching Programs (the 111 Project) under Grant B07048, and in part by the Major Research Plan of the National Natural Science Foundation of China under Grant 91438201 and Grant 91438103.

ABSTRACT Semi-supervised learning has been successfully connected in the research fields of machine learning such as data mining and dynamic data analysis. Imbalance class learning is one of the most challenging issues for classification. In recent years, the core focal point of numerous researchers has been on data classification of multi-class imbalanced datasets. In this paper, we proposed semi-supervised deep Fuzzy C-mean clustering for imbalanced multi-class classification (DFCM-MC). In our paper, the word ''Deep'' is used to show how decomposition strategy is applied deeply, first, decomposes the original semi-supervised data into supervised (labeled) and unsupervised (unlabeled) data. For training the model, we used unlabeled data along with labeled data to extract discriminative information, which is useful for classification. Second, it further decomposes the supervised and unsupervised data into multi intra-cluster that to address the problem of multi-class imbalance data, which tends to maximize intra-cluster classes and intra-cluster features. We propose a novel approach DFCM-MC by utilizing multi-intra clusters to extract new features to control redundancy for multi-class imbalance classification, which associates the maximum similarity of features between multi-intra clusters. Furthermore, we improve the classification performance of the DFCM-MC, apply the re-sampling technique to handle the imbalance data for classification. We conduct our experiments on 18 benchmark multi-class imbalanced datasets to demonstrate the performance of our proposed approach with the four state-of-the-art learning algorithms for multi-class imbalance data with three performance measures (mean of accuracy, mean of f-measure, and mean of area under the curve). The experiment results demonstrate that our proposed approach performs better due to their capacity to recognize and consolidate fundamental information from unsupervised data.

INDEX TERMS Semi-supervised learning, imbalanced data, multi-class classification, Fuzzy C-mean clustering, and feature learning.

I. INTRODUCTION

Real world applications have an abundance of data, but the challenges with data collection and labeling are expensive. However, in recent years researchers focused on making semi-supervised learning (SSL) structure in machine learning for improving accuracy with the large size of unlabeled data and the minimal size of labeled data. During the initial year of SSL, from the input labeled data are utilized to train the classifier during the training process and then using out-of-sample approaches to predict the labels for unlabeled data. Now, recently SSL approaches fall in sequential learning of supervised and unsupervised learning with the highest confidence score, until the convergence this procedure is repeated [1]–[4]. However, a couple of researchers have been utilized simultaneously supervised and unsupervised data to extract the useful information from unsupervised data to supervised data [5]–[8]. Many SSL approaches are utilized for classification. However, most of them focused on balanced classes [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

Class imbalance is generally referred to the situation, where the number of samples from one class is much higher or lower than that from other classes. Class imbalance is a general issue encountered in machine learning for most traditional classification problems, in which few classes are exceptionally in minority when compared to other classes [10]. This is a challenging situation when trying to classify the minority class, however, minority class is constantly more of interest. In recent studies, there are many different types of approaches to handle the problem of class imbalance datasets, such as random sampling, subset approaches, and cost-sensitive learning algorithms which combine more than one approach. There are many well-known learning algorithms for binary class imbalance data, KNN [11], [12], SMOTE [13], Easy-Ensemble, and Balance Cascade [14] etc. In our approach, we address both majority and minority class samples during data pre-processing technique. Random under-sampling (RUS) or Random over-sampling (ROS) is used to balance the number of features between the majority and minority classes.

Nevertheless, the class imbalanced problem is not the only factor that affect the performance of the prediction model. High dimensionality datasets also affect the performance prediction due to high computational cost. Few feature reduction based classification model are proposed [15] by eliminating the irrelevant features.

In our approach, we focus on eliminating the redundant features and handling the problem of multi-class imbalance data for classification [10]. Previously, many researchers have been focused on binary imbalanced learning for classification. However, Multi-class imbalanced data is encountered in real-world applications [16]–[18], the problem of overlapping between the distinct classes in datasets make a more challenging problem than binary imbalanced learning. In the recent study, to address the multi-class imbalanced data issue for classification either adopt ensemble-based approaches [19], [20], [21], [71] decomposition strategies [22], in which the multi-class imbalanced data is divided into several binary class subsets, which is easy-to-solve.

In this paper, we design a new approach for multi-class imbalanced data classification, namely DFCM-MC, which is the extension of our previous work [23]–[25]. We extend our previous work for the binary imbalanced dataset to the multi-class imbalanced dataset by utilizing decomposition strategy on two layers. The first layer decomposes semi-supervised data into supervised and unsupervised, which simultaneously operate during the training process, in which the user information is extracted from unlabeled data to support the development of a good classifier. In the second layer, the supervised data is further decomposed into subsets accordingly to the number of classes for (one-vs.-one) deep relationship among supervised and unsupervised data. However, to the best of our knowledge, very few works are done on multi-clusters to overcome the issue of the class imbalanced problem [26].

Feature learning is a crucial process for realizing embedded information in data analysis. By transforming data into low dimension for efficient learning [27]. The classification performance highly depends on the features, which is used as input to design the classifier. Generally believed, more features are redundant, irrelevant causing more risk by making the system complex and furthermore growths the time and cost. Hence, generally feature can be reduced into two ways; feature extraction [28], [29], and feature selection [30], [31]. However, very few works have been done on combine feature reduction technique to enhance the performance of classification on the multi-class imbalanced dataset.

Re-sampling is a method to balance the number of samples between the groups or classes of the datasets, which is widely used in data pre-processing [10], [13], [14]. Random under-sampling (RUS) and Random over-sampling (ROS) [31] are the two main techniques for resampling. RUS is used to reduce the data from majority classes and ROS is used to increase the data in minority class; both are helpful for a class imbalance problem and easy to implement with better results.

Our approach DFCM-MC used semi-supervised data in which, unsupervised data (unlabeled data) is used with their predicted labels using FCM clustering. For classification, we used labeled data (supervised data) and unlabeled data (unsupervised data) with their predicted labels which extract the discriminative information which is used for classification.

The motivation to utilize the combine feature reduction techniques to handle the problem of imbalanced data and also eliminate the irrelevant and redundant features and noisy data for classification by using proposed DFCM-MC based feature extraction technique and feature selection technique (Random under-sampling (RUS) and Random over-sampling (ROS)) [31].

We summarized the contribution of our approach as follows.

1. In this paper, we focus on multi-class imbalanced datasets for classification. The "Deep" word in our paper is utilized for deeply decompose semi-supervised data into multi-clusters to address the multi-class imbalanced issue.
2. We propose a novel approach DFCM-MC based feature extraction technique to deal with multi-class imbalance dataset to redundancy control for classification, which associates the maximum similarity between the intra-cluster classes (within the cluster of classes) and intra-cluster features (within the cluster of features) by using FCM clustering.
3. In order to enhance the prediction ability, we design feature extraction technique with random sampling to handle the problem of imbalanced data.

The remainder paper is organized as follows. In section 2, we briefly review of recent advances work on multi-class imbalanced data classification, Section 3, we introduce our new proposed DFCM-MC algorithm, Section 4, we describe the experimental results and analysis, section 5, provides with the threats to validity and section 6, finally conclude the paper.

II. RELATED WORK

Semi-supervised learning (SSL) is an active research area in machine learning. Many researchers have used SSL for binary and multi-class classification techniques [5], [32]–[41].

Ao *et al.* [5] proposed unconstrained probabilistic embedding by combining supervised and unsupervised models, in their approach, to improve the classification accuracy of the supervised model in which the ensemble learning is used to the output from supervised and unsupervised models. However, all these semi-supervised classification methods are based on balanced classes and they cannot handle the problem of overlapping.

The FCM algorithm [32], [42] based on the distance between samples, initially, the selecting of centers is complex due to the faults. In practical problem, FCM may be confined into local optimal. To increase grouping efficiency and solving problems with proximity issues, the semi-supervised FCM clustering approach may be a better choice for an imbalanced class problem. Many researchers have proposed Semi-supervised FCM based algorithms for classification and clustering [32], [34], [43].

The data pre-processing is valuable to enhance the classification performance and decrease the time cost [44]–[46], which includes feature reduction and resampling techniques. Feature reduction is used to increase the generalization performance of classification [15], [47]–[53] by removing the irrelevant features from the balanced and imbalanced datasets. However, all these methods are focused on binary imbalance problem.

In recent study, multi-class imbalance learning techniques either based on decomposition strategies [19], [37] and ensemble-based approaches [10], [22], [54]. Decomposition strategy is used to deal with the multi-class imbalance data by dividing the more complicated original problems into several easier-to-solve binary class sub-sets [55].

Sáez *et al.* [56], analyzing the overlapping between the distinct classes in multi-class datasets, they studied two methods, AdaBoost.NC [38] and Static-SMOTE [19]. AdaBoost.NC with random over-sampling is a representative method with negative correlation learning and adding punishment parameter when weighting the sample to encourage ensemble diversity. Static-SMOTE with resampling technique for "r" steps in the data pre-processing phase, where r is the number of classes. In each iteration, the resampling strategy initially chooses the class on the bases of minimum size and then add the same number of instances as present in the original dataset by applying the SMOTE algorithm.

Hoens [57], proposed an improved decision tree technique for multi-class imbalance datasets by using Hellinger distance and decomposition strategy as the splitting criterion. Fernández [58], proposed classification technique for multi-class imbalance dataset by using decomposition scheme (i.e. one versus one and one versus all) on both pre-processing of data and cost-sensitive learning with respect to several ad hoc approaches. They can find the good behavior achieved by the synergy between pairwise learning and resampling learning. Many other decomposition strategy based classification algorithms for multi-class imbalance dataset are proposed [41], [59]–[61].

Vluymans [41], proposed dynamic affinity-based technique for multi-class imbalance data classification by using decomposition strategy (one-versus-one) with a fuzzy rough set (FROVOCO) approach by locating of adaptive weight for binary classifier, addressing the unpredictable characteristic of the sub-problems and develop a novel dynamic aggregation technique for classification of the binary classifiers with the global class affinity making a final conclusion.

García [54], proposed Dynamic ensemble selection technique for multi-class imbalanced datasets, they propose a weighting mechanism to enhance the capability of classifier that is more powerful in the region of imbalanced datasets.

However, analysis of the previous research, most of it to utilized multi-cluster to handle the class imbalance problem [26] and does not take into consideration. Therefore, if we can apply the multi-cluster approach to select the most suitable features according to each class, the performance of the classification may be better.

In this paper, we propose multi-cluster based feature extraction technique by decomposition strategy using FCM clustering on intra-clusters to select the most appropriate features and apply under-sampling for imbalanced dataset for classification of multi-class imbalanced datasets.

III. SEMI-SUPERVISED DEEP FUZZY C - MEAN CLUSTERING FOR IMBALANCED MULTI-CLASS CLASSIFICATION (DFCM-MC)

In this paper, we present a Semi-Supervised DFCM clustering for imbalanced multi-class classification algorithm which provides an effective and practical approach by using decomposition strategies for predicting labels for unlabeled data that fall into a particular class.

We proposed semi-supervised DFCM-MC based feature extraction technique for classification to deal with imbalanced multi-class learning problems include ambiguity, small disjuncts, noisy data, and multi-class overlapping. Decomposition strategy not only handles the semi-supervised learning problem, but also deal with problems of ambiguity, and small disjuncts. DFCM-MC based feature extraction approach to deal with the noisy data, and multi-class overlapping problems. The integration of random sampling with our proposed feature extraction approach improves the performance of imbalanced multi-class classification.

A. FEATURE EXTRACTION

Utilization of many features customarily increases the data acquisition time and costs, it leads to more design time, more decision-making time, and some point it might result in more

Algorithm 1 DFCM-MC Membership and Centroid

Input:

The dataset $X = \{x_1, x_2 \ldots, x_n\}$, with n data points, r classes and k is features (clusters), the objective threshold is, fuzziness $m = 2$ and t is the number of iterations.

$X = X_{1L} \cup X_{2L} \cup \ldots \cup X_{rL} \cup X_{UNL} = X_{iL} \cup X_{UNL}$, where $i = 1, 2, \ldots, r$

$X_{iL} = \left\{x_{n(i-1)+1}, x_{n(i-1)+2}, \ldots, x_{n(i-1)+n_2}\right\} \in i^{th}$ class,

$X_{UNL} = \{x_{l+1}, x_{l+2} \ldots, x_n\} \in Unlabeled$ class,

Where, n_i is the number of data points in i^{th} class.

$\sum_{i=1}^{r} n_i = n_1 + n_2 + \ldots n_r = l$, where l is the total number of labeled data points.

Output:

Membership matrices U_{iL} *and* $U_{UNL}, \forall i \in 1, 2, \ldots r$

Set of k centroid $V_{iL}^{(k)}$ *and* $V_{UNL}^{(k)}, \forall i \in 1, 2, \ldots r$

1. Construct membership matrices U_{iL} & U_{UNL} by assigning membership of each data with random decimal fraction.
2. Calculate the sets of cluster center $V_{iL}^{(k)}$ & $V_{UNL}^{(k)}$ by using the formula of cluster center of FCM [62].
3. Update U_{iL} & U_{UNL} by using the formula of membership of FCM [32].
4. Repeat step 2 & 3 until $\left\|\mathrm{J}^{(t)} - \mathrm{J}^{(t-1)}\right\| < \varepsilon$ for all $r + 1$ labeled and unlabeled datasets separately.

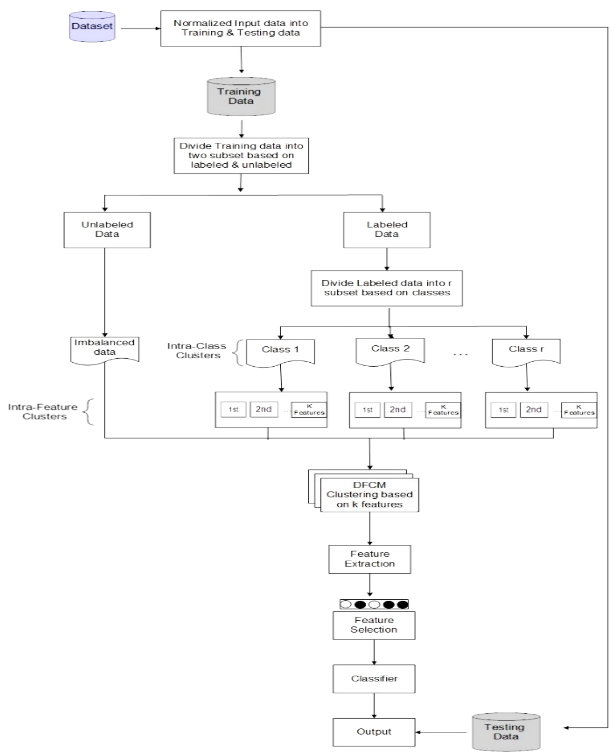

FIGURE 1. Flowchart of semi-supervised deep fuzzy C-mean clustering for imbalanced multi-class (DFCM-MC) classification.

risk. In this way, it is constantly attractive for classification that the number of features reduces the accumulated the decision-making system. There are two main fundamental approaches to decrease the feature dimensions i.e. feature extraction [28], [29], and feature selection [31], [63].

Algorithm 2 DFCM-MC Based Feature Extraction

Input:

The dataset $X = \{x_1, x_2 \ldots, x_n\}$ of n data points, r classes, and k features.

Sets of centroid $V_{iL}^{(k)}$ & $V_{UNL}^{(k)}$

Output:

$f_{iLq}(x_{iL})$ & f_{UNLq} $\forall$ $i \in 1, 2, \ldots, r$, sets of extracted features of i^{th} labeled class and unlabeled dataset. Where q is the number of extracted features.

1. Calculate Z_{iLk} & Z_{UNLk}, using formula
   Where,
   $Z_{iLk} = \left\|x_{iL} - V_{iL}^{(k)}\right\|, \quad \forall i \in 1.2, \ldots, r$ &
   $Z_{UNLk} = \left\|x_{UNL} - V_{UNL}^{(k)}\right\|$.
2. Calculate $\mu_{iL}(Z_{iL})$ & $\mu_{UNL}(Z_{UNL})$ are the means of the elements Z_{iL} & Z_{UNL} $\forall$ $i = 1, 2, \ldots, r$.
3. $f_k(x) = \max(0, \mu(z) - z_k) \,\forall\, iL$ & UNL *features*
4. Update all the features of all $r + 1$ subsets f_{iLq} & f_{UNLq}.

For the better performance of our approach, we utilized both approaches to design good prediction system, which deals with noisy data by eliminating redundant and irrelevant features and imbalanced multi-class problems.

For DFCM-MC based feature extraction in algorithm 2, we used FCM clustering to learn k centroid from multi-clusters based on features of labeled and unlabeled subsets. Given the DFCM-MC centroids $V^{(k)}$ by using algorithm 1. We choose non-linear mapping for feature mapping.

$$f_k(x) = \max(0, \mu(z) - z_k) \tag{1}$$

where $z_k = \left\|x - V^{(k)}\right\|$ and $\mu(z)$ is the mean of elements of z. If the output 0 of any feature f_k, where the distance to the centroid $V^{(k)}$ is "above average". In practice, this means that roughly half of the feature will be set 0, shown in table 8.

Random-sampling is utilized for feature selection to balance the number of extracted features between all $(r + 1)$ labeled and unlabeled subsets. Select "$s = s_1$", which is the selected number of features from each subsets by reducing the number of features from majority groups to equal to the number of features in minority groups by using RUS and "$s = s_2$", which is selected number of features from each subsets by increasing the number of features from minority groups to equal to the number of features in majority group by using ROS.

B. FINAL DFCM-MC CLASSIFICATION

The final classification by algorithm 3 is based on the maximum similarity between the features of unlabeled data points and labeled classes. Euclidean distance is chosen to measure the similarity between "s" cluster centers of labeled classes

TABLE 1. Description of dataset.

Dataset	No. Of Samples	No. Of Features	No. of Classes	Distribution of Class	MaxIR	MinIR	AvgIR
Automobile	150	25	6	3/20/48/46/29/13	16	1.04	4.9
Balance	625	4	3	288/49/288	5.88	1.00	4.25
Cleveland	297	13	5	164/55/36/35/13	12.62	1.03	3.87
Contraceptive	1473	9	3	629/333/511	1.89	1.23	1.55
Dermatology	358	34	6	111/60/71/48/48/20	5.55	1.00	2.17
Ecoli	336	7	8	143/77/2/2/35/20/5/52	71.5	1.00	15.27
Glass	214	9	6	70/76/17/13/9/29	8.44	1.31	3.60
Led7digit	500	7	10	45/37/51/57/52/52/47/57/53/49	1.50	1.00	1.16
New-Thyroid	215	5	3	150/35/30	5.00	1.09	3.48
Page-Blocks	5472	10	5	4913/329/28/87/115	175.46	1.32	31.65
Shuttle	58000	9	7	45586/49/171/8903/3267/10/13	4558.6	1.30	561.92
Thyroid Disease	7200	21	3	166/368/6666	40.16	2.22	20.16
Wine	178	13	3	59/71/48	1.48	1.20	1.30
Wine Quality-Red	1599	11	6	10/53/681/638/199/18	68.10	1.07	18.83
Yeast	1484	8	10	244/429/463/44/51/163/35/30/20/5	92.6	11.65	11.65
Lymphography	148	18	4	2/81/61/4	40.5	1.33	18.30
Car Evaluation	1728	6	4	384/69/1210/65	18.62	1.06	8.64
Zoo	101	17	7	41/20/5/13/4/8/10	10.25	1.25	3.12

Algorithm 3 DFCM-MC Classifier

Input:
The dataset $X = \{x_{l+1}, x_{l+2} \ldots, x_n\}$, with s selected features, $V_{iL}^{(s)}$ & $V_{UNL}^{(s)}$, $\forall i \in 1, 2, \ldots, r$.
Output:
Predicted labeled data $Y = \{y_{l+1}, y_{l+2} \ldots, y_n\}$

1. $Y = \emptyset$
2. **For** $j = l + 1$ **to** n & $i \in 1$ *to* r **do**
   Computing $max\ Sim_i \left(x_j, V_L^{(s)}\right)$ by using equation 2.
3. **If** max_$avg_max\ Sim_i \left(x_j, V_L^{(s)}\right) \in i^{th}$ labeled class
4. Adding x_j into i^{th} class
5. Updating all rest data points in X into Y
6. **return**Y.

and unlabeled data points with "s" features.

$$max\ Sim_i \left(x_j, V_L^{(s)}\right) = min \left|x_j - V_{iL}^{(s)}\right|, \quad \forall i = 1, 2, \ldots, r \tag{2}$$

where $x_j \in X_{UNL}$, V_{iL} is the set of the centroid of i^{th} labeled class and "s" is the number of selected features by using random-sampling from $(r + 1)$ subsets. With each selected "s" feature clusters, find the one to one maximum similarity between the features of unlabeled data and labeled classes. In the final classification stage, find the maximum average of the maximum similarity between the "s" selected features of unlabeled data and r labeled classes one to one. Then, the unlabeled data point comparing toward the maximum average of the maximum similarity including in the specific labeled class.

IV. EXPERIMENT

In this section, we establish the details of our experimental study on which we demonstrate the performance of our proposed semi-supervised classification approach on imbalanced multi-class datasets.

A. DATA PREPARATION

In our paper, we utilized MATLAB 2018a [65] as the programming tool. We demonstrate the performance of our proposed DFCM-MC algorithm on eighteen multi-classes imbalanced UCI datasets [66] with 10%, 20%, and 30% rate of labeled data. Table 1 demonstrates the benchmark UCI datasets that show brief properties of eighteen multi-class imbalanced datasets that incorporate the number of sample size, number of classes, number of features, distribution of class, and imbalanced ratio.

B. PERFORMANCE MEASURE

There are several ways for evaluating the performance of imbalanced multi-class classification calculate the performance of our proposed approach, we used three performance measure to estimate the performance of our approach by Mean Accuracy (MAcc), Mean of F-Measure (MFM), and Mean of Area Under the Curve (MAUC) [67].

The MAcc is obtained by the average value of the accuracy rate of each class independently. MAcc is defined as.

$$MAcc = \frac{\sum_{i=1}^{n} MAcc_i}{r} \tag{3}$$

where r is the number of classes and Acc_i is the average rate for i^{th} class.

TABLE 2. Confusion matrix.

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	True Positive (TP)	False Negative (FN)
Actual Negative Class	False Positive (FP)	True Negative (TN)

Table 2 shown the confusion matrix for binary class data. Here, majority class is considered negative and minority class is considered positive.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$F - Measure = \frac{2.recall.Precision}{recall + Precision} \tag{6}$$

For multi-class data, Mean F-Measure can be defined as follows.

$$MFM = \frac{\sum_{i=1}^{r} (FM)_i}{r} \tag{7}$$

where r is the total number of classes and i is the index for positive class.

Mean AUC is the average of the pairwise AUC values of all pairs of classes which is defined as.

$$\begin{aligned} MAUC &= \frac{2}{r(r-1)} \sum_{i<j} (AUC(C_i, C_j) \\ &= \frac{2}{r(r-1)} \sum_{i<j} \left[A\left(C_i, C_j\right) + A(C_j, C_i)\right] \end{aligned} \tag{8}$$

For two classes C_i & C_j, the value of AUC (C_i, C_j) represents the probability of being assigned to the i^{th} class by the classifier. When a randomly selected sample from the first class (i^{th} class) has a higher probability to assign compared to a randomly selected sample from the second class (j^{th} class) and vice versa.

C. EXPERIMENTAL SETUP

To evaluate the performance of our approach, we design our experiments on 18 benchmarks multi-class UCI datasets [32] with 10%, 20%, & 30% rate of labeled data by using three well-known performance measure (MAcc, MFM, & MAUC) for multi-class.

The details of the proposed method are already described in section 3. The experimental comparison is divided into two parts; (1) first, we compare our proposed approach with other state-of-the-art methods for imbalanced multi-class datasets to investigate the stability and efficiency of our approach, (2) second, we conduct an internal comparison of our proposed approach, in order to show the effect of the performance of RUS and ROS with DFCM-MC feature extraction technique to construct new dataset for classification. So we should choose the stable random-sampling technique (RUS or ROS), which is no account of the information loss in the training process.

We investigate our proposed approach by using only RUS for feature selection with other four imbalanced multi-class classification methods i.e. AdaBoost.NC [10], Static-SMOTE [19], FROVOCO (Fuzzy Rough OVO Combination) (FR) [41], and Dynamic Ensemble selection for multi-class imbalanced datasets (DES-ML) [54] on eighteen UCI datasets. We select m=2 as the degree of fuzziness and objective threshold 0.1 to stop the iteration for updating new multi-clusters. All final results are on the average of 100 runs.

D. STATISTICAL TESTS

For statistical analysis, we utilized two non-parametric statistical tests, e.g. for the pairwise comparison, we used Wilcoxon's signed-rank test [68] and for multiple comparisons, we used the Friedman test [69] with Holm post-hoc procedure [70]. Wilcoxon's signed-rank test is the comparison between the two methods according to the ranks. The smallest and the largest absolute difference in the result of two methods is assigned as rank "1" and rank "0", where D is the number of observations (datasets) eighteen datasets in our study. R^+ and R^- are the sum of ranks of positive and negative differences. When the p-value of Wilcoxon's signed-rank test [68] is smaller than the significance level of $\alpha = 0.05$, we can say that a significant difference is performed between the two methods. The Friedman test [69] is also ranked in combination with Holm post-hoc procedure [70]. The null hypothesis of the Friedman test is that all methods under consideration perform equivalently. When it is rejected, the post-hoc procedure is applied to detect where the significant differences can be found. The Friedman test is based on the ranking procedure and the lowest rank of any method showed the overall best performance. The Holm post-hoc procedure is utilized to compare all other methods, for this purpose, the lowest rank is used as a base method. When the p-value is smaller than the significance level α, we can say that the base method outperforms the other methods. We denoted the P_{Friedman} p-value of the Friedman test and (P_{Holm}) is adjusted p-value of the post-hoc procedure.

E. RESULTS & ANALYSIS

In this section, we analysis our results in three parts; (1) first, compare our results with state-of-the-art in terms of performance measure (MAcc, MFM, and MAUC), and also statistical analysis, (2) Second, analysis the effect of random sampling technique in our proposed approach, and (3) third, analysis the effectiveness of our proposed approach on imbalanced ratio (IR).

1) COMPARISON OF DFCM-MC WITH STATE-OF-THE-ART METHODS

To demonstrate the effectiveness of our proposed approach on the imbalanced multi-class dataset, we compare our approach with four well known state-of-the-art methods AdaBoost.NC [10], Static-SMOTE [19], FROVOCO (Fuzzy Rough OVO Combination) (FR) [41], and Dynamic Ensemble selection for multi-class imbalanced datasets (DES-ML) [54] for imbalanced multi-class classification.

Table 3, 4 & 5 shows all the results of classification methods on 18 benchmark imbalanced multi-class datasets with three performance measures (MAcc, MFM, & MAUC) respectively. The best performance for each dataset is highlighted in bold. We observe from table 3, 4 & 5, our proposed approach performs better for 15 out of 18 datasets

TABLE 3. MAcc of DFCM-MC with compared methods using 18 UCI datasets on average of labeled rate (0.1, 0.2, & 0.3).

Dataset	Ada-NC	Static-SMOTE	FR	DES - MI	DFCM - MC
Automobile	69.73	74.80	77.16	86.30	**89.13**
Balance	62.86	57.87	78.85	63.89	**79.14**
Cleveland	31.60	28.19	33.78	30.55	**35.67**
Contraceptive	**53.53**	47.01	52.72	52.99	52.67
Dermatology	95.55	92.70	97.15	96.00	**99.37**
Ecoli	68.19	65.80	79.27	72.69	**78.02**
Glass	71.57	70.79	67.07	76.48	**80.47**
Led7digit	42.34	70.87	64.79	73.47	**73.37**
New-Thyroid	90.79	90.62	91.11	93.87	**95.61**
Page-Blocks	73.42	81.71	90.04	77.76	**93.26**
Shuttle	91.78	90.26	91.85	91.80	**92.35**
Thyroid Disease	**97.31**	92.36	66.45	95.94	94.18
Wine	89.24	93.02	98.21	96.18	**98.66**
Wine Quality-Red	39.44	33.33	43.75	42.89	**45.21**
Yeast	54.14	50.08	58.82	58.26	**60.39**
Lymphography	88.01	78.41	86.24	84.74	**88.02**
Car Evaluation	**96.09**	94.07	92.31	95.56	95.36
Zoo	87.74	89.05	90.88	91.79	**94.07**
Avg	72.41	72.27	75.58	76.73	**80.28**

TABLE 4. MFM of DFCM-MC with compared methods using 18 UCI datasets on average of labeled rate (0.1, 0.2, & 0.3).

Dataset	Ada-NC	Static-SMOTE	FR	DES - MI	DFCM - MC
Automobile	0.6245	0.7309	0.7501	0.8438	**0.9026**
Balance	0.6098	0.5774	0.7300	0.6274	**0.8107**
Cleveland	0.3027	0.2726	0.3027	0.2945	**0.3232**
Contraceptive	0.5221	0.4701	0.4805	0.5027	**0.5226**
Dermatology	0.9526	0.9236	0.9350	0.9568	**0.9638**
Ecoli	0.5958	0.6029	0.7287	0.6574	**0.7355**
Glass	0.6286	0.6855	0.6601	0.7401	**0.7532**
Led7digit	0.3674	0.7045	0.6374	0.7284	**0.7291**
New-Thyroid	0.8970	0.9006	0.9374	**0.9419**	0.9301
Page-Blocks	0.6875	0.7632	0.8721	0.7300	**0.8921**
Shuttle	0.8714	0.8777	0.9101	0.9074	**0.9136**
Thyroid Disease	**0.9374**	0.9357	0.6503	0.9315	0.9296
Wine	0.8894	0.9289	0.9358	0.9606	**0.9621**
Wine Quality-Red	0.3079	0.3268	0.4096	0.4011	**0.4236**
Yeast	0.4609	0.4930	0.5600	0.5550	**0.5885**
Lymphography	0.7907	0.7163	0.8117	0.7904	**0.8297**
Car Evaluation	**0.9481**	0.9302	0.9470	0.9472	0.9463
Zoo	0.8427	0.8597	0.8505	0.8961	**0.9133**
Avg	0.6798	0.7055	0.7283	0.7451	**0.7816**

for both MAcc and MFM performance measures, and 14 out of 18 datasets for MAUC performance measure when compared with compared methods. Secondly, DFCM-MC, DES-MI, and FR are based on the data pre-processing method, which proposed to generate a balanced training dataset. DES-MI and FR are addressed the biases toward majority class samples and ignore the minority class samples, which are not consistent with their true labels. Our approach DFCM-MC address both majority and minority class samples during data pre-processing stage, which are consistent with their true labels and extract only discriminative information which is useful for classification.

In figure 2, 3 & 4, the proposed approach lead the performance on all methods when compare the average on 18 datasets. We cannot extract any meaningful conclusions without the statistical analysis. Therefore, we used Wilcoxon's test for the pairwise comparison between DFCM-MC and other state-of-the-art methods for three performance measures, and Friedman test with Holm post-hoc procedure for the comparison among all the imbalanced multi-class classification methods for three performance measures.

TABLE 5. MAUC of DFCM-MC with compared methods using 18 UCI datasets on average of labeled rate (0.1, 0.2, & 0.3).

Dataset	Ada-NC	Static-SMOTE	FR	DES - MI	DFCM - MC
Automobile	0.8370	0.8699	0.9633	0.9681	**0.9732**
Balance	0.8309	0.6909	0.8854	0.8421	**0.8961**
Cleveland	0.6334	0.5972	0.6981	0.6123	**0.7213**
Contraceptive	0.6667	0.6160	0.6485	0.6621	**0.6813**
Dermatology	0.9856	0.9461	0.9966	0.9891	**0.9954**
Ecoli	0.7362	0.7290	0.9304	0.8521	**0.9531**
Glass	0.9146	0.8704	0.9325	0.9416	**0.9613**
Led7digit	0.7040	0.9134	0.9189	0.9421	**0.9503**
New-Thyroid	0.9572	0.9463	0.9981	0.9903	**0.9983**
Page-Blocks	0.8076	0.8939	0.9736	0.8325	**0.9762**
Shuttle	0.9211	0.9579	**0.9987**	0.9923	0.9981
Thyroid Disease	**0.9698**	0.9365	0.8494	0.9393	0.9383
Wine	0.9218	0.9588	**1.0000**	0.9938	0.9931
Wine Quality-Red	0.7581	0.7495	0.8342	0.8526	**0.8753**
Yeast	0.8279	0.8372	0.8810	0.8629	**0.8866**
Lymphography	0.9347	0.8317	0.9108	0.8916	**0.9531**
Car Evaluation	**0.9965**	0.9672	0.9563	0.9901	0.9889
Zoo	0.8854	0.9013	0.9276	0.9416	**0.9711**
Avg	0.8494	0.8452	0.9057	0.8943	**0.9284**

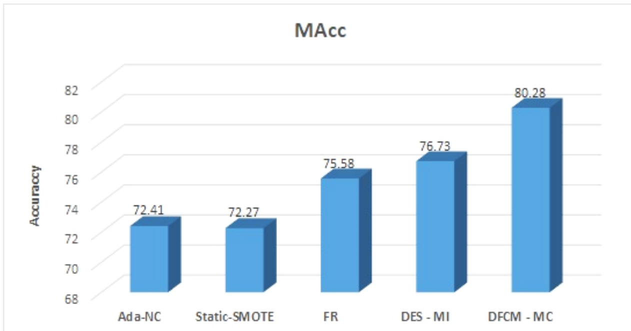

FIGURE 2. Comparison of MAcc of DFCM-MC other compared methods.

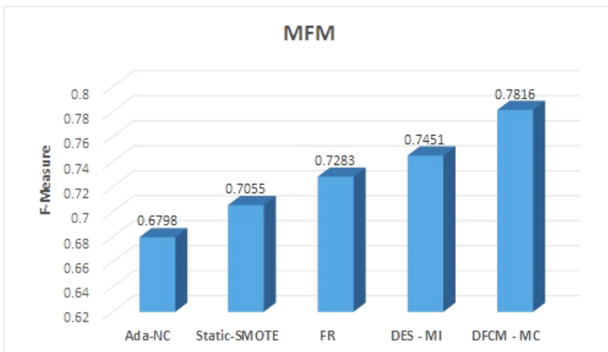

FIGURE 3. Comparison of MFM of DFCM-MC other compared methods.

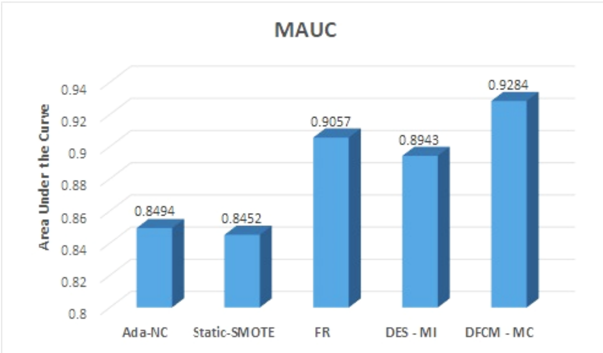

FIGURE 4. Comparison of MAUC of DFCM-MC other compared methods.

TABLE 6. Friedman test and Holm post hoc procedure on three performance measures (MAcc, MFM, & MAUC).

Performance Measure	Methods	Rank		P_{Holm}
MAcc	DFCM-MC	1.556	(1)	-
	Ada-NC	3.556	(4)	**0.000913**
	Static-SMOTE	4.444	(5)	**0.00000318**
	FR	2.833	(3)	0.084101
	DES-MI	2.667	(2)	0.17452
	P_{Friedman}	**0.000001033**		
MFM	DFCM-MC	1.444	(1)	-
	Ada-NC	3.9722	(4)	**0.000016**
	Static-SMOTE	4.000	(5)	**0.000012**
	FR	2.9167	(3)	**0.041653**
	DES-MI	2.667	(2)	0.138765
	P_{Friedman}	**0.000002**		
MAUC	DFCM-MC	1.456	(1)	-
	Ada-NC	3.778	(4)	**0.00093**
	Static-SMOTE	4.444	(5)	**0.0000012**
	FR	2.611	(2)	0.174542
	DES-MI	2.722	(3)	**0.01878**
	P_{Friedman}	**0.0000000907**		

Based on the statistical results in table 6 and 7, we can observe from Table 6, our proposed approach is assigned the lowest rank for all performance measures. The p-value of Friedman and Holm post-hoc procedure for all performance measures are less than 0.05 is highlighted in bold. DFCM-MC significantly outperforms all other methods except DES-MI for (MAcc and MFM) and FR for (MAcc and MAUC) performance measures. It does not show that our method is not significantly better then DES-MI and FR, although as stated above the calculated Friedman rank shows that our method is best.

Also for the deep comparison in table 7, our method achieved the most wins performance in datasets when compared with other compared methods pairwise. For MAcc, DFCM-MC wins 15 out of 18, 18 out of 18, 16 out of 18, and 15 out of 18 datasets when compared to Ada-NC, Static-SMOTE, FR, and DES-MI respectively, for MFM, 16 out of 18, 17 out of 18, 16 out of 18, and 15 out of 18 datasets when compared to Ada-NC, Static-SMOTE, FR, and DES-MI respectively, and for MAUC, 16 out of 18, 18 out of 18, 15 out of 18, and 15 out of 18 datasets when compared to Ada-NC, Static-SMOTE, FR, and DES-MI respectively. The all p-values of Wilcoxon's test for all performance measure are less than 0.05, our method significantly outperforms all other methods. Based on all the above analysis, our method performed better for imbalance multi-class datasets when compared with other imbalanced multi-class classification methods. It shows that our approaches of decomposition technique and multi-clustering based feature extraction technique can boost the classification performance for multi-class datasets.

2) ANALYSIS OF THE EFFECT OF RANDOM SAMPLING TECHNIQUE IN OUR PROPOSED APPROACH

For the analysis of the effect of random sampling techniques in our proposed approach, we use Random undersampling (RUS), and Random oversampling (ROS) [64] for imbalanced datasets. The random sampling is used to balance the data after the decomposition of semi-supervised data and DFCM-ML based feature extraction. Table 8, shows the details of all feature extraction in $(\mathbf{r}+\mathbf{1})$ subsets and feature selection by using RUS and ROS. $\mathbf{s_1}$ is the number of selected features by ROS, and $\mathbf{s_2}$ is the number of selected features by RUS. From the results of table 8, 52% features are reduced by RUS, and 22% features are reduced by ROS. RUS better due to fewer features for computational time.

When we will do the comparison of RUS and ROS for classification performance, we calculate the results of our proposed approach with RUS and ROS on 18 benchmark datasets with the labeled rate 10%, 20% and 30% for three performance measure (MAcc, MFM, and MAUC). From table 9, 10, and 11, we conclude that DFCM-RUS perform better than DFCM-ROS on all performance measures with all labeled rates. By statistical analysis, we used Wilcoxon's test on the average of all labeled rates (10%, 20%, & 30%) for all three performance measures.

Table 12, shows the results of Wilcoxon's test for MAcc, MFM, and MAUC, it shows that our proposed approach using RUS (DFCM-RUS) wins 18 out of 18 datasets, 15 out of 18 datasets, and 15 out of 18 datasets for MAcc, MFM, and MAUC performance measure respectively, when compared to DFCM-ROS. All p-values of Wilcoxon's test are less than 0.05, which shows that DFCM-RUS significantly outperform DFCM-ROS. Due to better performance of the proposed approach with RUS, for the final results in table 3, 4, & 5, we used RUS in our proposed approach.

3) ANALYSIS OF THE EFFECTIVENESS OF OUR PROPOSED APPROACH ON IMBALANCED RATION (IR)

For the analysis of the effectiveness of our proposed method on IR, we used the results of tables 3, 4 & 5 with the maximum, minimum, and average IR information from table 1. It can be observed that the performance of our method on the

TABLE 7. Wilcoxon's test for pairwise comparison between DFCM-MC and state-of-the-art methods for three performance measures (MAcc, MFM, & MAUC).

Performance Measure	Methods	W / L	R^+	R^-	$P_{wilcoxon}$	Significant (Yes / No)
MAcc	DFCM-MC vs. Ada-NC	15 / 3	159	12	0.00138	Yes
	DFCM-MC vs. Static-SMOTE	18 / 0	171	0	0.00020	Yes
	DFCM-MC vs. FR	16 / 2	165	6	0.00054	Yes
	DFCM-MC vs. DES-MI	15 / 3	159	12	0.00138	Yes
MFM	DFCM-MC vs. Ada-NC	16 / 2	166	5	0.00044	Yes
	DFCM-MC vs. Static-SMOTE	17 / 1	170	1	0.00024	Yes
	DFCM-MC vs. FR	16 / 2	166	5	0.00044	Yes
	DFCM-MC vs. DES-MI	15 / 3	158	13	0.00158	Yes
MAUC	DFCM-MC vs. Ada-NC	16/ 2	165	6	0.00054	Yes
	DFCM-MC vs. Static-SMOTE	18 / 0	171	0	0.00020	Yes
	DFCM-MC vs. FR	15 / 3	160	11	0.00120	Yes
	DFCM-MC vs. DES-MI	15 / 3	165	6	0.00054	Yes

TABLE 8. Details of selected features by feature extraction algorithm with (RUS and ROS).

Dataset	No. of feature extraction in all subsets (r subset of labeled classes & 1 subset of unlabeled classes) $(C_1, C_2, \ldots, C_r, unlabeled)$	ROS		RUS	
		s_1 (No. of selected features by ROS)	Percentage of selected features	s_2 (No. of selected features by RUS)	Percentage of selected features
Automobile	14, 16, 50, 18, 17, 15, 16	20	0.80	14	0.56
Balance	3, 2, 3, 3	3	0.75	2	0.50
Cleveland	10, 8, 7, 7, 6, 9	10	0.77	6	0.46
Contraceptive	7, 4, 6, 7	7	0.78	4	0.44
Dermatology	25, 18, 16, 15, 14, 14, 23	25	0.74	14	0.41
Ecoli	5, 5, 4, 4, 5, 4, 4, 5, 5	5	0.71	4	0.57
Glass	6, 7, 4, 4, 4, 5, 6	7	0.78	4	0.44
Led7digit	3, 3, 4, 6, 4, 4, 3, 6, 5, 5, 6	6	0.86	3	0.43
New-Thyroid	3, 2, 2, 3	3	0.60	2	0.40
Page-Blocks	8, 6, 5, 5, 6, 7	8	0.80	5	0.50
Shuttle	7, 5, 6, 7, 6, 4, 4, 6	7	0.78	4	0.44
Thyroid Disease	10, 11, 17, 16	17	0.81	10	0.48
Wine	8, 10, 6, 10	10	0.77	6	0.46
Wine Quality-Red	6, 7, 9, 9, 8, 6, 8	9	0.82	6	0.55
Yeast	6, 7, 7, 5, 5, 6, 5, 5, 5, 4, 7	7	0.88	4	0.50
Lymphography	8, 14, 12, 9, 13	14	0.78	8	0.44
Car Evaluation	4, 3, 5, 3, 5	5	0.83	3	0.50
Zoo	13, 12, 9, 10, 9, 10, 10, 12	13	0.76	9	0.53
Avg			0.78		0.48

top six MaxIR datasets (Ecoli, Page-Block, Shuttle, Thyroid Disease, Wine Quality-Red, and Lymphography). The performance of our methods wins on 5 out of 6 MaxIR datasets for MAcc & MFM, and 4 out of 6 MaxIR datasets for MAUC. Similarly, we can observe that the results for the top six MinIR datasets. According to table 3, 4, & 5, our method is also outperformed on six MinIR datasets for all performance measure. We can conclude the results from the above analysis, our method not only achieves the good performance on MinIR datasets, and also our proposed method boost the classification performance on imbalance multi-class dataset with MaxIR.

V. THREATS TO VALIDITY

Our experimental results are affected by some threats of validation.

TABLE 9. MAcc results of proposed approach with ROS and RUS with labeled rate (0.1, 0.2, & 0.3).

DATASET	Labeled Rate = 0.1		Labeled Rate = 0.2		Labeled Rate = 0.3	
	DFCM-ROS	DFCM-RUS	DFCM-ROS	DFCM-RUS	DFCM-ROS	DFCM-RUS
Automobile	79.21	**85.10**	81.43	**88.70**	86.26	**93.59**
Balance	71.43	**74.97**	73.43	**78.27**	76.69	**84.18**
Cleveland	28.29	**33.65**	31.63	**36.39**	32.21	**36.97**
Contraceptive	50.17	**52.13**	51.19	**53.38**	50.07	**52.50**
Dermatology	94.37	**98.97**	95.61	**99.40**	95.97	**99.74**
Ecoli	69.27	**71.79**	70.63	**79.31**	75.76	**82.96**
Glass	**80.13**	80.11	**80.46**	80.07	80.71	**81.23**
Led7digit	66.27	**72.16**	69.96	**73.56**	72.16	**74.39**
New-Thyroid	92.28	**93.98**	95.78	**96.02**	95.81	**96.83**
Page-Blocks	93.46	**93.79**	93.52	**93.26**	92.61	**92.73**
Shuttle	91.26	**92.16**	91.67	**92.76**	90.78	**92.13**
Thyroid Disease	88.76	**90.59**	89.47	**95.23**	92.36	**96.72**
Wine	97.63	**98.16**	97.79	**98.59**	98.21	**99.23**
Wine Quality-Red	34.73	**41.37**	36.21	**46.65**	41.21	**47.61**
Yeast	56.21	**59.97**	56.29	**61.38**	56.31	**59.82**
Lymphography	80.76	**88.01**	83.37	**88.01**	90.21	**88.04**
Car Evaluation	91.26	**94.97**	91.68	**95.27**	92.21	**95.84**
Zoo	86.67	**91.67**	90.26	**94.21**	90.46	**96.33**
Avg	75.12	**78.53**	76.69	**80.58**	78.33	**81.71**

TABLE 10. MFM results of proposed approach with ROS and RUS with labeled rate (0.1, 0.2, & 0.3).

DATASET	Labeled Rate = 0.1		Labeled Rate = 0.2		Labeled Rate = 0.3	
	DFCM-ROS	DFCM-RUS	DFCM-ROS	DFCM-RUS	DFCM-ROS	DFCM-RUS
Automobile	0.7678	**0.8778**	0.7921	**0.9016**	0.8026	**0.9284**
Balance	0.7517	**0.8096**	0.7531	**0.8099**	0.7531	**0.8126**
Cleveland	0.2997	**0.3301**	0.3067	**0.3206**	0.3170	**0.3189**
Contraceptive	0.5121	**0.5215**	0.5171	**0.5225**	0.5191	**0.5289**
Dermatology	0.9017	**0.9321**	0.9117	**0.9796**	0.9121	**0.9797**
Ecoli	0.6716	**0.7198**	0.6778	**0.7200**	0.7091	**0.7667**
Glass	0.6541	**0.6927**	0.7049	**0.7768**	0.7063	**0.7901**
Led7digit	0.6397	**0.7297**	0.6949	**0.7310**	0.6853	**0.7266**
New-Thyroid	0.9121	**0.9263**	0.9141	**0.9317**	0.9227	**0.9323**
Page-Blocks	0.8338	**0.8860**	0.8491	**0.8920**	0.8573	**0.8983**
Shuttle	0.7873	**0.8579**	0.8321	**0.9067**	0.8556	**0.9762**
Thyroid Disease	0.9011	**0.9268**	0.9026	**0.9295**	0.9036	**0.9325**
Wine	0.8711	**0.9001**	0.9003	**0.9931**	0.9007	**0.9931**
Wine Quality-Red	0.4117	**0.4230**	0.4210	**0.4233**	0.4210	**0.4245**
Yeast	0.5437	**0.5526**	0.5446	**0.5536**	0.6013	**0.6593**
Lymphography	0.8127	**0.8232**	0.8297	**0.8313**	**0.8357**	0.8346
Car Evaluation	0.9321	**0.9451**	0.9333	**0.9460**	0.9356	**0.9478**
Zoo	0.8541	**0.8969**	0.8565	**0.9026**	0.8678	**0.9404**
Avg	0.7255	**0.7640**	0.7412	**0.7818**	0.7503	**0.7995**

A. CONSTRUCT VALIDITY

The new datasets after pre-processing data for classification are the main threat to the construct validity. We take no account to address the distributional difference between original dataset and newly constructed dataset. Our proposed approach is designed to convert the original multi-class imbalanced dataset into several small subsets by decomposition strategy and resampling is to avoid the effect of imbalance data, the final results of our experiment are the average of 100 runs. Further threats to the construct validity are that whether the performance measure used in our experiments are suitable for the multi-class imbalanced dataset. We choose Mean of Accuracy (MAcc), Mean of F-Measure (MFM), and Mean of Area under the curve (MAUC) [67] for the multi-class imbalanced dataset to evaluate the performance for classification for the multi-class imbalanced dataset.

B. INTERNAL VALIDITY

The selection of parameters might be the threats to the internal validity. We choose FCM clustering to design our proposed DFCM-MC algorithm, which is widely used in machine learning. Considering that many researchers is used to taking default parameters in empirical studies, we also take the default parameters of the FCM clustering model in WEKA. To avoid the other threat to internal validity, all the execution is cross-checked by our lab research group. Consequently, we believe there are negligible threats are remaining.

C. EXTERNAL VALIDITY

The quality of the input data (original or constructed) for the classification model may be the most important threat to external validity. To avoid this threat, our experimental results

TABLE 11. MAUC results of proposed approach with ROS and RUS with labeled rate (0.1, 0.2, & 0.3).

DATASET	Labeled Rate = 0.1		Labeled Rate = 0.2		Labeled Rate = 0.3	
	DFCM-ROS	DFCM-RUS	DFCM-ROS	DFCM-RUS	DFCM-ROS	DFCM-RUS
Automobile	0.8510	**0.9710**	0.8727	**0.9727**	0.8834	**0.9759**
Balance	0.8733	**0.8964**	0.8771	**0.8966**	0.8771	**0.8953**
Cleveland	0.5996	**0.6978**	0.6924	**0.7147**	0.7121	**0.7514**
Contraceptive	0.6717	**0.6801**	0.6810	**0.6812**	0.6817	**0.6826**
Dermatology	0.9421	**0.9927**	0.9476	**0.9961**	0.9527	**0.9974**
Ecoli	0.8873	**0.9499**	0.9062	**0.9541**	0.9078	**0.9553**
Glass	0.9573	**0.9615**	0.9526	**0.9614**	0.9601	**0.9610**
Led7digit	**0.9481**	0.9471	0.9416	**0.9497**	0.9401	**0.9541**
New-Thyroid	0.9717	**0.9982**	0.9817	**0.9983**	0.9808	**0.9984**
Page-Blocks	0.9124	**0.9667**	0.9427	**0.9801**	0.9026	**0.9818**
Shuttle	0.9817	**0.9979**	0.9937	**0.9987**	0.9938	**0.9977**
Thyroid Disease	0.8023	**0.8913**	0.8710	**0.9603**	0.9011	**0.9633**
Wine	0.9914	**0.9935**	0.9915	**0.9932**	0.9915	**0.9926**
Wine Quality-Red	0.7827	**0.8535**	0.8217	**0.8723**	0.8397	**0.9001**
Yeast	0.8556	**0.8719**	0.8667	**0.8797**	0.8811	**0.9082**
Lymphography	0.9027	**0.9467**	0.9476	**0.9631**	0.9363	**0.9495**
Car Evaluation	0.9787	**0.9878**	0.9823	**0.9888**	0.9847	**0.9901**
Zoo	0.9123	**0.9578**	0.9441	**0.9779**	0.9494	**0.9776**
Avg	0.8790	**0.9201**	0.9008	**0.9299**	0.9042	**0.9351**

TABLE 12. Wilcoxon's test for pairwise comparison between DFCM-RUS and DFCM-ROS for three performance measure.

Performance Measure	Methods	W / L, D	R^+	R^-	$P_{wilcoxon}$	Significant (Yes / No)
MAcc	DFCM-RUS vs. DFCM-ROS	18 / 0, 0	171	0	0.00020	Yes
MFM	DFCM-RUS vs. DFCM-ROS	15 / 0, 3	120	0	0.00064	Yes
MAUC	DFCM-RUS vs. DFCM-ROS	15 / 2, 1	143.5	9.5	0.00152	Yes

are produced based on benchmark multi-class imbalanced datasets from UCI [66] datasets which are commonly used for the classification of multi-class imbalanced datasets. Also, we select FCM clustering as a based method for feature extraction and re-sampling (RUS and ROS) for feature selection, which is broadly utilized in imbalanced classification, which is ensured to converge.

D. STATISTICAL VALIDITY

For the statistical analysis, we performed two non-parametric statistical tests to evaluate the significance of the differences in performance. For multiple comparisons, whether significant differences exist within a group of methods, we used Friedman test [69] in combination with Holm post-hoc procedure [70] for multi-class classification.

VI. CONCLUSION

In our paper, we provided deep decomposition approach, semi-supervised Deep FCM clustering for multi-class imbalanced classification, which incorporate DFCM-MC feature extraction technique after decomposition strategy and resampling to amend the quality of datasets for the multi-class imbalance classification model. The decomposition process is utilized for split the semi-supervised data into supervised and unsupervised data, which simultaneously deal during the training phase to extract the useful information from unsupervised data to supervised data to support the development of good classifier which is further decomposed into multi-cluster that tends to maximize intra-cluster classes and intra-cluster features to deal with multi-class problem. FCM is utilized for the multi-clustering for feature extraction to create new upright features and also eliminate redundant and irrelevant features and noisy data for input data for classification. Resampling is utilized to deal with imbalance dataset.

In our experiment, in the first part, we compared our approach with four state-of-the-art multi-class imbalance learning algorithms. The highest average of all performance measure MAcc, MFM, and MAUC of proposed approach show the potential of our method compared to all other methods. In the second part, internal comparison of our approach, in order to show the effect of resampling (RUS & ROS) with DFCM-MC feature extraction for classification performance. Our experiment results show the potential of our approach with RUS in improving the imbalance multi-class classification performance on 18 benchmark datasets with all three performance measure (MAcc, MFM, and MAUC). For statistical analysis Wilcoxon signed-rank-test for pairwise comparison and Friedman with Holm post-hoc procedure for comparison among the multi-class imbalance algorithms. The statistical result demonstrates that DFCM-MC method is significantly better than other compared methods.

In our future work, we attempt to improve the effectiveness of our proposed approach by using a spark-based approach. We will also check the stability of our approach to image datasets.

REFERENCES

[1] S. Melacci and M. Belkin, "Laplacian support vector machines trained in the primal," *J. Mach. Learn. Res.*, vol. 12, pp. 1149–1184, Jul. 2011.

[2] Z. Qi, Y. Tian, and Y. Shi, "Laplacian twin support vector machine for semi-supervised classification," *Neural Netw.*, vol. 35, no. 11, pp. 46–53, Nov. 2012.
[3] K. P. Bennett and A. Demiriz, "Semi-supervised support vector machines," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 11, 1999, pp. 368–374.
[4] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proc. Int. Conf. Mach. Learn.*, 1999, pp. 200–209.
[5] X. Ao *et al.*, "Combining supervised and unsupervised models via unconstrained probabilistic embedding," *Inf. Sci.*, vol. 257, pp. 101–114, Feb. 2014.
[6] S. Basu, A. Banerjee, and R. J. Mooney, "Semi-supervised clustering by seeding," in *Proc. 19th Int. Conf. Mach. Learn.*, 2002, pp. 27–34.
[7] W. Cai, S. Chen, and D. Zhang, "A simultaneous learning framework for clustering and classification," *Pattern Recognit.*, vol. 42, no. 7, pp. 1248–1259, 2009.
[8] N. V. Chawla and G. J. Karakoulas, "Learning from labeled and unlabeled data an empirical study across techniques and domains," *J. Artif. Intell. Res.*, vol. 23, pp. 331–366, Mar. 2005.
[9] S. Lessmann, B. Baesens, C. Mues, and S. Pietsch, "Benchmarking classification models for software defect prediction: A proposed framework and novel findings," *IEEE Trans. Softw. Eng.*, vol. 34, no. 4, pp. 485–496, Jul. 2008.
[10] S. Wang and X. Yao, "Multiclass imbalance problems: Analysis and potential solutions," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 42, no. 4, pp. 1119–1130, Aug. 2012.
[11] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York, NY, USA: Wiley, 1973.
[12] S. Vluymans, D. S. Tarragó, Y. Saeys, C. Cornelis, and F. Herrera, "Fuzzy rough classifiers for class imbalanced multi-instance data," *Pattern Recognit.*, vol. 53, pp. 36–45, May 2016.
[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jan. 2002.
[14] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," in *Proc. Int. Conf. Data Mining (ICDM)*, 2006, pp. 539–550.
[15] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proc. 14th Int. Conf. Artif. Intell. Statist. (AISTATS)*, vol. 15, 2011, pp. 215–223.
[16] H. Yu, S. Hong, X. Yang, J. Ni, Y. Dan, and B. Qin, "Recognition of multiple imbalanced cancer types based on DNA microarray data using ensemble classifiers," *BioMed Res. Int.*, vol. 2013, Jul. 2013, Art. no. 239628.
[17] X. M. Zhao, X. Li, L. Chen, and K. Aihara, "Protein classification with imbalanced data," *Proteins, Struct., Function, Bioinf.*, vol. 70, no. 4, pp. 1125–1132, 2008.
[18] S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery: A small target detection benchmark," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 187–203, Jan. 2016.
[19] G. Haixiang *et al.*, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017.
[20] S. Riaz, A. Arshad, and L. Jiao, "Fuzzy rough C-mean based unsupervised CNN clustering for large-scale image data," *Appl. Sci.*, vol. 8, p. 1869, Oct. 2018.
[21] S. Riaz, A. Arshad, and L. Jiao, "Rough noise-filtered easy ensemble for software fault prediction," *IEEE Access*, vol. 6, pp. 46886–46899, 2018.
[22] F. Fernández-Navarro, C. Hervás-Martínez, and P. A. Gutiérrez, "A dynamic over-sampling procedure based on sensitivity for multi-class problems," *Pattern Recognit.*, vol. 44, no. 8, pp. 1821–1833, 2011.
[23] A. Arshad, S. Riaz, L. Jiao, and A. Murthy, "A semi-supervised deep fuzzy C-mean clustering for two classes classification," in *Proc. IEEE 3rd Inf. Technol. Mechatron. Eng. Conf. (ITOEC)*, Chongqing, China, Oct. 2017, pp. 365–370.
[24] A. Arshad, S. Riaz, L. Jiao, and A. Murthy, "Semi-supervised deep fuzzy C-mean clustering for software fault prediction," *IEEE Access*, vol. 6, pp. 25675–25685, 2018.
[25] A. Arshad, S. Riaz, L. Jiao, and A. Murthy, "The empirical study of semi-supervised deep fuzzy c-mean clustering for software fault prediction," *IEEE Access*, vol. 6, pp. 47047–47061, 2018.
[26] G. Forestier and C. Wemmert, "Semi-supervised learning using multiple clusterings with limited labeled data," *Inf. Sci.*, vols. 361–362, pp. 48–65, Sep. 2016.
[27] L. Maaten, E. Postma, and J. Herik, "Dimensionality reduction: A comparative review," Tech. Rep. TiCC TR 2009-005, TiCC, Tilburg Univ., Tilburg, The Netherlands, 2009.
[28] N. R. Pal and V. K. Eluri, "Two efficient connectionist schemes for structure preserving dimensionality reduction," *IEEE Trans. Neural Netw.*, vol. 9, no. 6, pp. 1142–1154, Nov. 1998.
[29] N. R. Pal, V. K. Eluri, and G. K. Mandal, "Fuzzy logic approaches to structure preserving dimensionality reduction," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 3, pp. 277–286, Jun. 2002.
[30] N. R. Pal and K. K. Chintalapudi, "A connectionist system for feature selection," *Neural Parallel Sci. Comput.*, vol. 5, no. 3, pp. 359–382, 1997.
[31] D. Chakraborty and N. R. Pal, "A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 110–123, Jan. 2004.
[32] K. Li, Z. Cao, L. Cao, and R. Zhao, "A novel semi-supervised fuzzy c-means clustering method," in *Proc. Chin. Control Decis. Conf.*, Guilin, China, Jun. 2009, pp. 3761–3765.
[33] S. Basu, M. Bilenko, and R. J. Mooney, "Comparing and unifying search-based and similarity-based approaches to semi-supervised clustering," in *Proc. Workshop Continuum Labeled Unlabeled Data Mach. Learn. Data Mining (ICML)*, 2003, pp. 42–49.
[34] C. Zhao, H. Zhao, and W. Yao, "Fuzzy C-means clustering based on improved marked watershed transformation," *Telkomnika*, vol. 14, no. 3, pp. 981–986, 2016.
[35] N. Grira, M. Crucianu, and N. Boujemaa, "Active semi-supervised fuzzy clustering," *Pattern Recognit.*, vol. 41, pp. 1834–1844, May 2008.
[36] Y.-F. Li, J. T. Kwok, and Z.-H. Zhou, "Semi-supervised learning using label mean," in *Proc. 26th Annu. Int. Conf. Mach. Learn.* Montreal, QC, Canada, 2009, pp. 633–640.
[37] J. Bi and C. Zhang, "An empirical comparison on state-of-the-art multi-class imbalance learning algorithms and a new diversified ensemble learning scheme," *Knowl.-Based Syst.*, vol. 158, pp. 81–93, Oct. 2018.
[38] S. Wang, H. Chen, and X. Yao, "Negative correlation learning for classification ensembles," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2010, pp. 2893–2900.
[39] A. S. Ghanem, S. Venkatesh, and G. West, "Multi-class pattern classification in imbalanced data," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2881–2884.
[40] X.-Y. Liu, Q.-Q. Li, and Z.-H. Zhou, "Learning imbalanced multi-class data with optimal dichotomy weights," in *Proc. IEEE 13th Int. Conf. Data Mining (IEEE ICDM)*, Dec. 2013, pp. 478–487.
[41] S. Vluymans, A. Fernández, Y. Saeys, and C. Cornelis, "Dynamic affinity-based classification of multi-class imbalanced data with one-versus-one decomposition: A fuzzy rough set approach," *Knowl. Inf. Syst.*, vol. 56, pp. 55–84, Jul. 2018.
[42] D. S. Mai and L. T. Ngo, "Semi-supervised fuzzy C-means clustering for change detection from multispectral satellite image," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Istanbul, Turkey, Aug. 2015, pp. 1–8.
[43] H. Husain, M. Khalid, and R. Yusof, "Automatic clustering of generalized regression neural network by similarity index based fuzzy C-means clustering," in *Proc. IEEE Region 10 Conf. (TENCON)*, vol. 2, Nov. 2004, pp. 302–305.
[44] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[45] K. Gao, T. M. Khoshgoftaar, H. Wang, and N. Seliya, "Choosing software metrics for defect prediction: An investigation on feature selection techniques," *Softw.-Practice Exper.*, vol. 41, no. 5, pp. 579–606, 2011.
[46] S. Shivaji, E. J. Whitehead, R. Akella, and S. Kim, "Reducing features to improve code change-based bug prediction," *IEEE Trans. Softw. Eng.*, vol. 39, no. 4, pp. 552–569, Apr. 2013.
[47] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, Oct. 2004.
[48] W. Liu, S. Liu, Q. Gu, J. Chen, X. Chen, and D. Chen, "Empirical studies of a two-stage data preprocessing approach for software fault prediction," *IEEE Trans. Rel.*, vol. 65, no. 1, pp. 38–53, Mar. 2016.
[49] H. Wang, T. M. Khoshgoftaar, and A. Napolitano, "An empirical study on the stability of feature selection for imbalanced software engineering data," in *Proc. 11th Int. Conf. Mach. Learn. Appl.*, Boca Raton, FL, USA, vol. 1, Dec. 2012, pp. 317–323.
[50] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal.*, vol. 6, no. 5, pp. 429–449, Oct. 2002.

[51] T. G. Grbac, G. Mausa, and B. D. Basic, "Stability of software defect prediction in relation to levels of data imbalance," in *Proc. 2nd Workshops Softw. Qual. Anal., Monit., Improvement, Appl.*, Novi Sad, Serbia, Sep. 2013, pp. 1–10.

[52] D. Ryu, O. Choi, and J. Baik, "Value-cognitive boosting with a support vector machine for cross-project defect prediction," *Empirical Softw. Eng.*, vol. 21, no. 1, pp. 43–71, 2016.

[53] Q. Yu, S. Jiang, And Y. Zhang, "The performance stability of defect prediction models with class imbalance: An empirical study," *IEICE Trans. Inf. Syst.*, vol. E100-D, no. 2, pp. 265–272, Feb. 2017.

[54] S. García, Z.-L. Zhang, A. Altalhi, S. Alshomrani, and F. Herrera, "Dynamic ensemble selection for multi-class imbalanced datasets," *Inf. Sci.*, vols. 445–446, pp. 22–37, Jun. 2018.

[55] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "NMC: Nearest matrix classification–A new combination model for pruning One-vs-One ensembles by transforming the aggregation problem," *Inf. Fusion*, vol. 36, pp. 26–51, Jul. 2017.

[56] J. A. Sáez, B. Krawczyk, and M. Woźniak, "Analyzing the oversampling of different classes and types of examples in multi-class imbalanced datasets," *Pattern Recognit.*, vol. 57, pp. 164–178, Sep. 2016.

[57] T. Hoens, Q. Qian, N. Chawla, and Z.-H. Zhou, "Building decision trees for the multi-class imbalance problem," in *Prco. 16th Pacific–Asia Conf. Knowl. Discovery Data Mining*, in Lecture Notes in Computer Science, vol. 7301. Kuala Lumpur, Malaysia: Springer, May/Jun. 2012, pp. 122–134.

[58] A. Fernández, V. López, M. Galar, M. J. del Jesus, and F. Herrera, "Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches," *Knowl.-Based Syst.*, vol. 42, pp. 97–110, Apr. 2013.

[59] Z. Zhang, B. Krawczyk, S. Garcìa, A. Rosales-Pérez, and F. Herrera, "Empowering one-vs-one decomposition with ensemble learning for multi-class imbalanced data," *Knowl.-Based Syst.*, vol. 106, pp. 251–263, Aug. 2016.

[60] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognit.*, vol. 44, no. 8, pp. 1761–1776, 2011.

[61] A. C. Lorena, A. C. P. L. F. de Carvalho, and J. M. P. Gama, "A review on the combination of binary classifiers in multiclass problems," *Artif. Intell. Rev.*, vol. 30, nos. 1–4, pp. 19–37, 2008.

[62] K. Pal, R. K. Mudi, and N. R. Pal, "A new scheme for fuzzy rule-based system identification and its application to self-tuning fuzzy controllers," *IEEE Trans. Syst., Man, Cybern. B, (Cybern.)*, vol. 32, no. 4, pp. 470–482, Aug. 2002.

[63] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-18, no. 1, pp. 183–190, Jan./Feb. 1988.

[64] T. M. Khoshgoftaar, C. Seiffert, J. Van Hulse, A. Napolitano, and A. Folleco, "Learning with limited minority class data," in *Proc. Int. Conf. Mach. Learn. Appl.*, Dec. 2007, pp. 348–353.

[65] *Statistics Toolbox Release 2018a*, MATLAB, The MathWorks, Natick, MA, USA, Mar. 2018.

[66] *UCI Repository of Machine Learning Databases*. Accessed: Aug. 2018. [Online]. Available: http://www.ics. uci.edu/mlearn/MLRepository.html

[67] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Mach. Learn.*, vol. 45, no. 2, pp. 171–186, 2001.

[68] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.* vol. 1, no. 6, pp. 80–83, 1945.

[69] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Stat. Assoc.*, vol. 32, no. 200, pp. 675–701, 1937.

[70] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, no. 2, pp. 65–70, 1979.

[71] S. Riaz, A. Arshad, and L. Jiao, "Rough-KNN noise-filtered convolutional neural network for image classification," in *Proc. 3rd Int. Conf. Inf. Technol. Intell. Transp. Syst. (ITITS)*, vol. 314, Xi An, China, pp. 265–275, 2018.

ALI ARSHAD received the B.S. degree in computer science from Iqra University, Pakistan, in 2008, and the M.S. degree in software engineering from International Islamic University, Pakistan, in 2012. He is currenlty pursuing the Ph.D. degree with the School of Computer Science and Technology, Xidian University, China. His research interests include machine learning, semi-supervised learning, and Fuzzy C mean clustering.

SAMAN RIAZ received the M.Phil. and M.Sc. degrees in applied mathematics from Quaid-e-Azam, Pakistan, in 2008 and 2006, respectively. She is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Xidian University, China. Her research interests include machine learning, Fuzzy rough C-mean, and probability.

LICHENG JIAO received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively. Since 1992, he has been a Professor with the School of Artificial Intelligence, Xidian University, Xi'an, where he is currently the Director of the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education of China. He has published more than 20 monographs and 100 papers in international journals and conferences. He is in charge of about 40 important scientific research projects. His research interests include image processing, natural computation, machine learning, and intelligent information processing. He is a Fellow of the IEEE Xi'an Section Execution Committee, the Chairman of the Awards and Recognition Committee, the Vice Board Chairperson of the Chinese Association of Artificial Intelligence, the Councilor of the Chinese Institute of Electronics, a Committee Member of the Chinese Committee of Neural Networks, and an Expert of the Academic Degrees Committee of the State Council.