IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 26, 2020, accepted March 18, 2020, date of publication March 25, 2020, date of current version April 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2983317

Weighted Deep Forest for Schizophrenia Data Classification

YAFEI ZHU, SHUYUE FU, SHIHU YANG, PING LIANG, AND YING TAN
Key Laboratory for Computer Systems of State Ethnic Affairs Commission, Southwest Minzu University, Chengdu 610041, China

Corresponding author: Ying Tan (ty_edu@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC1706200, in part by the Sichuan Science and Technology Program under Grant 2019YFG0207 and Grant 2019YFH0055, and in part by the Fundamental Research Funds for the Central Universities, Southwest Minzu University, under Grant 2016NGJPY06.

ABSTRACT There is no objective biological indicator for the diagnosis of schizophrenia. Machine learning is used to classify functional magnetic resonance imaging (fMRI) data, the aim of which is to effectively improve the reliability of diagnostics for schizophrenia. The following points are often considered: 1) Extracting effective features from fMRI data. 2) Choosing an appropriate machine learning method. 3) Improving classification accuracy. In this paper, we propose a weighted deep forest model, which includes a weighted class vector, and a prediction class vector. In our experiment, we extract functional connection (FC) features from fMRI data. Then, we use principal component analysis (PCA) to reduce the dimension of FC features. For datasets with unbalanced data, we use SMOTE to balance the data. Finally, the datasets with balanced data are fed into the weighted forest model. Compared with the classification results obtained by traditional classifiers, our classification accuracy is better. This method will provide greater possibilities for assisting doctors in diagnosing schizophrenia. This paper has significance for the study of schizophrenia by helping doctors diagnose the disease.

INDEX TERMS Schizophrenia, fMRI, deep forest, classification.

I. INTRODUCTION

A. BACKGROUND

Schizophrenia is a severe psychiatric disorder that affects approximately 1% of people [1]. Patients often have thought disorders and negative symptoms. Moreover, inconsistent performance is evident in their mental activities and environment. Diagnosis of schizophrenia with high confidence is important in neurosciences and medical science [2], [3].

However, because the cause of schizophrenia is still poorly understood, the diagnosis of schizophrenia is mainly based on the patient's behavioral performance, which can be rated with tools such as the positive and negative syndrome scale (PANSS) [4]. It is extremely urgent to find an effective method to improve the diagnosis rate of schizophrenia.

Functional magnetic resonance imaging (fMRI) mainly relies on changes in blood oxygenation level-dependent (BOLD) signals caused by nerve activity; from these, the relevant activated brain area can be detected. The most prominent advantage of fMRI is that it enables the noninvasive detection of brain function in vivo and that the direct correlation between different brain regions can be observed at the same time.

As a neuroimaging technology, fMRI has been widely used in the diagnosis and analysis of schizophrenia [5]. Cheng *et al.* [6] obtained functional networks from fMRI data, and they then used a linear support vector machine (SVM) to classify patients with schizophrenia and members of the control group. The next year, Kim *et al.* [7] used a revised DNN to classify fMRI data, and they proposed an adaptive learning algorithm to explicitly control the weight sparsity in each hidden layer via L1-norm regularization. Yan *et al.* [8] also used the functional network connectivity as a feature; they used a deep neural network (DNN) and layer-wise relevance propagation (LRP) to classify schizophrenia. Zou *et al.* [9] built a 3-D convolutional neural networks (CNN) model to investigate the local spatial patterns of MRI features and designed a multi-modality CNN architecture to combine fMRI and sMRI features. Zhu *et al.* [10] used the parameter of asymmetry (PAS) and a support vector machine (SVM) to analyze an fMRI dataset in 2018, and they found disrupted asymmetry of inter- and intrahemispheric functional connectivity in patients. Qureshi *et al.* [11] use 3D-CNN to classify schizophrenia

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

from control. 3D ICA based functional connectivity networks were used as the input features of the classifier in their work.

In 2017, gcForest was proposed by Zhou and Feng [12]; it was a cascade forest model based on a deep neural network. It obtained new features through multigranularity scanning. The new features were input into a cascade layer to obtain classification results. Shao *et al.* [13] added data balancing in gcForest and used the revised gcForest to classify ADHD patients from control participants.

Although the research on schizophrenia using fMRI data is developing continuously, the methods proposed are still in the development stage. The level of these approaches is still not comparable to that of experts in the field of mental illness. The research on schizophrenia based on fMRI data still faces the following problems:

First, human understanding of brain structure and function is not deep enough [5]. The cause of schizophrenia and the abnormal brain pattern in patients with the disease is unclear. At present, the most commonly used feature of schizophrenia classification is to construct functional connections in the brain from fMRI data. Therefore, it is increasingly important to use mathematical tools to extract deep information from fMRI data.

Second, fMRI data are scarce. Sakai and Yamada [14] summarized that in 21 schizophrenia classification studies over the past five years (2014-2018), the average sample size was 208, and the median sample size was 147. Therefore, for such a small amount of data, choosing an appropriate machine learning method plays a key role.

B. OUR CONTRIBUTIONS

(1)A weighting method is used on the class vector in the traditional gcForest cascading layer, which makes the contribution of the class vector obtained by each validation different from the average class vector.

(2) We propose the softmax function to obtain the weight when we generate the prediction class vector at the last level. This makes the forest with higher training accuracy have a higher contribution.

(3) Principal component analysis (PCA) is applied to reduce the dimensionality of features, which aims to avoid the ''dimension curse'' problem.

(4) We use SMOTE to balance the data in the training set before the training set is fed into the multi-grained scanning layer, which avoids the problem of unbalanced data.

II. MATERIALS AND METHODS

A. DATASETS

We use the resting-state fMRI data from COBRE and UCLA. The COBRE (Center for Biomedical Research Excellence) dataset is available for download from the site (http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html). The UCLA dataset is available at the site (https://openfmri.org/).

The first dataset is COBRE, including 72 patients with schizophrenia and 75 healthy controls. The original fMRI data were obtained from a 3-Tesla SIEMENS TIM scanner with the following parameters: time of repetition TR=2000 ms, echo time TE=29 ms, flip angle FA=75°, and field of view FOV=192 mm with a 4 mm thickness and 0 mm gap; the matrix size is 64×64 and the number of axial slices is 32.

The second dataset is UCLA [15], [16], including 138 patients with schizophrenia and 58 healthy controls. The original fMRI data were obtained from a 3-Tesla SIEMENS TIM scanner with the following parameters: time of repetition TR=2000 ms, echo time TE=30 ms, flip angle FA=90°, and field of view FOV=192 mm, with a 4 mm thickness and 0 mm gap; the matrix size is 64×64 and the number of axial slices is 34.

B. DATA PROCESSING AND FEATURES GENERATION

All fMRI data were preprocessed by using a toolbox named DPABI [17], which can be freely downloaded from the web site (http://rfmri.org/dpabi). The preprocessing included slice time correction, realigning, normalization, and smoothing.

After the above processing steps, the image samples are 4-D, including 3-D spatial information and a 1-D time series. All brain images are divided into 116 brain regions according to the automated anatomical atlas (AAL) [18]. Then, the mean blood oxygen concentration of all brain regions is treated as the time series of all ROIs. The time series of all ROIs is 2-D; it includes spatial information and a 1-D time series.

For any pair in the time series, we calculate the Pearson correlation coefficient to form a functional connection (FC) matrix [19]. The calculation is as follows:

$$P_{ij} = \frac{\sum \left(S_i - \overline{S_i}\right)\left(S_j - \overline{S_j}\right)}{\sqrt{\sum_i \left(S_i - \overline{S_i}\right)^2}\sqrt{\sum_j \left(S_j - \overline{S_j}\right)^2}} \tag{1}$$

where P_{ij} is the Pearson correlation of time S_i and time S_j, and $\overline{S_i}$, $\overline{S_j}$ are the mean values of times S_i and S_j.

Then, we obtain a function connectivity matrix P.

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{n1} & \cdots & P_{nn} \end{bmatrix} \tag{2}$$

The connectivity matrix is a symmetric matrix, and the lower left triangular of the matrix forms a feature vector. By concatenating the first-row vector to the last-row vector of the lower left triangle, the features can be generated. The dimension of the FC feature vector is $((116*116\text{-}116)/2+116) = 6786$.

C. GCFOREST

A random forest, which is a tree-based ensemble machine technique, has advantages in dealing with nonlinear classification problems and overfitting [20]. Zhou and Feng [12] proposed a new tree-based ensemble method named gcForest. gcForest has two major structures, namely, multi-grained scanning and a cascade forest.

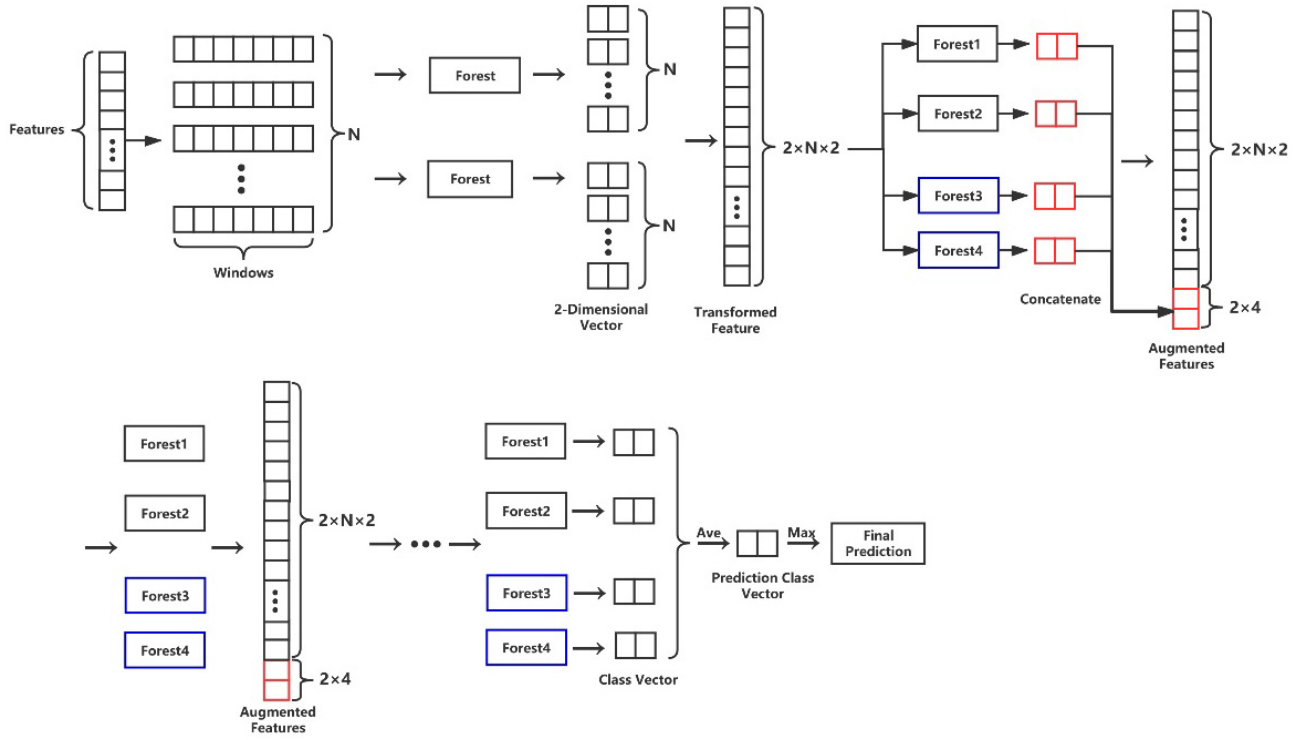


FIGURE 1. The overall procedure of gcForest.

Features of a given sample will be processed by multi-grained scanning, thus making N sets of features whose length is the size of the window. Then, each set is fed into two different forests to output a 2-dimensional vector as a class distribution vector. Furthermore, we concatenate all the output class vectors as one feature, called a transformed feature.

In a cascade forest, each level is an ensemble of decision tree forests. For an example with a transformed feature, each forest will produce a class distribution vector. Then, all the class distribution vectors generated by the forest at the same level are concatenated as augmented features, which are then concatenated with the transformed feature vector to be input to the next level of cascade. To avoid overfitting, a class vector is generated by k-fold cross-validation. In other words, each instance will be used as training data k-1 times, thus making k-1 class vectors. The final class vector is produced by averaging the k-1 class vectors.

In the last level, the prediction class vector is obtained by averaging four class vectors generated by four forests, and the largest probability is the prediction result.

D. REVISED METHODS

Cross-validation can be used to evaluate the predictive performance of models, especially the performance of trained models on new data, which can reduce overfitting to some extent. Therefore, we can use the accuracy of cross validation to judge the quality of a model and assign weights to the vectors based on the accuracy.

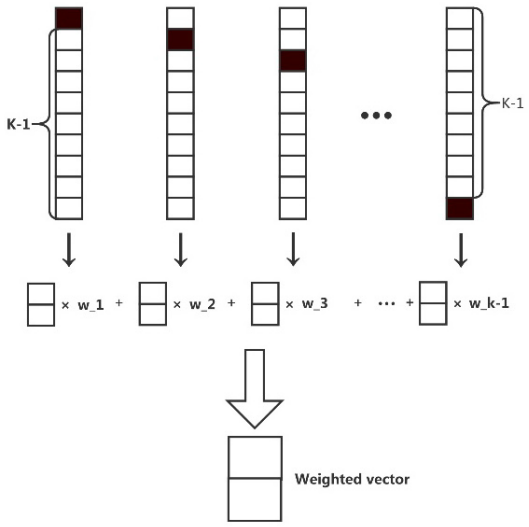


FIGURE 2. The generation of a weighted class vector.

1) WEIGHTED CLASS VECTOR OBTAINED BY THE FOREST

The class vector of an instance is generated by k-1 class vectors from k-fold cross-validation. We believe the contribution of each class vector of these k-1 class vectors is different because the model trained by each training set is different. Therefore, we use a weighting method to average these class vectors. Each weight depends on the test accuracy of the forest, and the method of assigning weights can be formulated as follows:

$$\widehat{v_J} = \sum_i w_{ij} v_{ij} \tag{3}$$

where:

$$w_i = \frac{acc_{ij}}{\sum_i acc_{ij}} \tag{4}$$

and where $\widehat{v_J}$ is the weighted class vector generated by the Jth forest, w_{ij} is the weight of the class vector obtained by the ith trained jth forest, v_{ij} the class vector obtained by the ith trained jth forest and acc_{ij} the test accuracy of the ith trained jth forest.

2) WEIGHTED PREDICTION VECTOR

The prediction result relies on the forests in the last level of the cascade. Each forest is different from the others. Furthermore, the contribution of each is different. Here, we propose that the contribution of a forest depends on the average accuracy of the forest and that the forest's weight is described by the softmax function. This proposal can be formulated as follows:

$$V = \sum_j W_J v_J \tag{5}$$

where:

$$W_J = f\left(\overline{ACC_J}\right) \tag{6}$$

$$f\left(\overline{ACC_J}\right) = \frac{\exp\left(\overline{ACC_J}\right)}{\sum_J \exp\left(\overline{ACC_J}\right)} \tag{7}$$

and where V is the prediction class vector from the last level of cascade, $\widehat{v_J}$ is the class vector from the Jth forest from the penultimate level of cascade, W_J is the weight of the class vector obtained by the Jth forest, $\overline{ACC_J}$ is the average test accuracy obtained by k-fold cross-validation of the Jth forest and f(x) is the softmax function.

Each x (average test accuracy) has a value between 0 and 1, and the output is between e^0 and e^1 when it goes through the softmax function. The independent variable is restricted to (0,1), thus making the strong values stronger and the weak weaker. The output of the stronger will not be much larger than the output of the weaker if the independent variable does not surpass 1. In other words, the stronger forests will not account for an enormous proportion when the class vector is averaged.

3) DIMENSIONALITY REDUCTION

Machine learning models have strong requirements on the dimensionality of data input. The "dimension curse" problem will arise when high-dimensional data are fed into a model to train the model parameters without feature extraction (e.g., in the case of fMRI data R^{n*p}, where n is the sample size, which is usually less than 500, and p is the number of features, which is usually more than 5000; thus, p is much larger than n). Therefore, we use PCA to reduce the data input dimension.

As a general method of data dimension reduction, PCA can ensure that the main information of data is extracted with minimum error. Its effectiveness has been proved in previous use [21]–[23]. It has the following two points: First, the principal components are orthogonal, which can eliminate the factors that affect the original data components. The second is to measure the amount of information in terms of variance, independent of factors other than the data set.

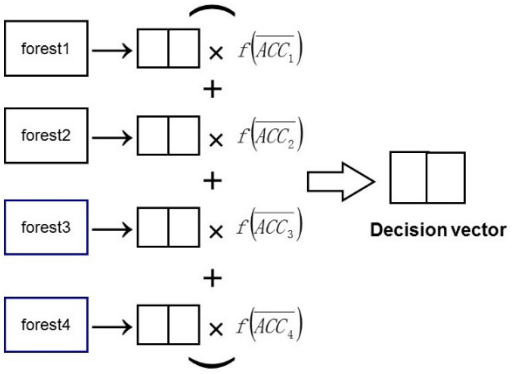


FIGURE 3. The generation of a weighted prediction vector.

4) DATA BALANCING

Since standard learning algorithm may generate suboptimal classifiers [24], it is necessary to deal with the imbalanced problem. An overfitting model will be trained when the sample size of control subjects in the data set is much larger than that of patients, and an overfit model is valueless as it will not predict any patients. Therefore, SMOTE [25] is used to increase the sample size of patients. However, because the sample size of control subjects is much larger than that of patients, direct use of SMOTE may add too much noise to the data. Therefore, we randomly cut the training set to reduce the difference between the sample size of patients and the sample size of control subjects. Then, SMOTE is used. In this way, the data balance is guaranteed without introducing too many noise points.

III. RESULT

A. EXPERIMENT

We combined the features (described in Section 2.2) and labels obtained from the Cobre and UCLA datasets to form the set of all samples. We used the Cobre dataset, the UCLA dataset and the combined Cobre and UCLA datasets to train and test the model. PCA was used to extract the principal components from FC features. When we selected the principal component, the 99% of the data information was selected after dimension reduction. In the Cobre dataset, the interpretation of PCA's parameter characteristics is shown in Figure 6.

The principal components were features with low dimension, which were fed into the model. K-Fold was used to divide the training set and test set. First, we randomly cut data with more categories out of the training set so that the labels of the training set tended to be balanced, and then we used SMOTE to balance the training set. We sent the balanced

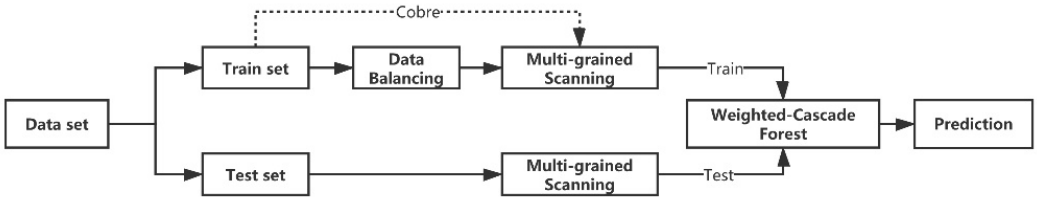


FIGURE 4. The experimental process of the revised gcForest.

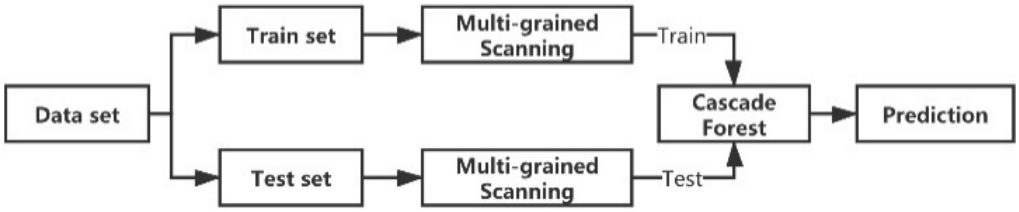


FIGURE 5. The experimental process of the traditional gcForest.

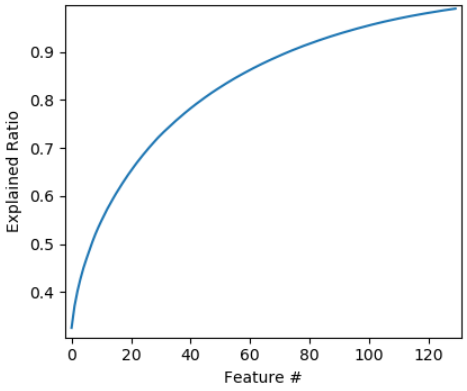


FIGURE 6. PCA component cumulated explained power.

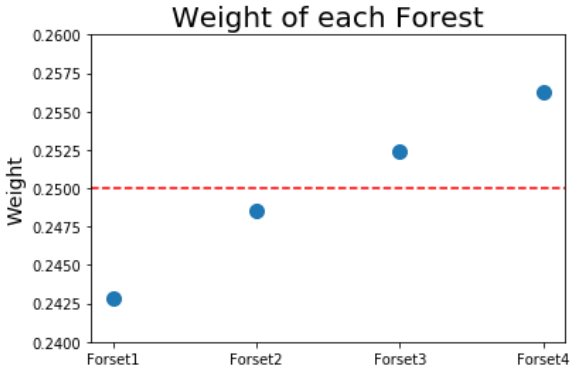


FIGURE 7. Weight of each forest in the revised gcForest.

training set to gcForest and weighted-gcForest for training models. Then, we made predictions on the test set and, finally, evaluated the model. It should be noted here that because the categories of the Cobre dataset are relatively balanced, we did not balance the Cobre data when we experimented with Cobre alone. The specific process is shown in Figure 4.

Then, we also experimented with a traditional forest. The experimental process of the traditional gcForest is shown in Figure 5.

In addition, SVM, a random forest and Linear Discriminant Analysis (LDA) were compared with the revised gcForest. For SVM, we used the radial basis function (RBF) SVM, and the penalty parameter misclassification cost weight C and kernel width γ were selected from $\{10^{-5},10^{-4},\ldots,10^{4},10^{5}\}$by the grid search method. For the random forest, the parameter n_estimators (the number of trees in the forest) was selected from $\{10^{1},\ldots,10^{4},10^{5}\}$by the grid search method.

Finally, we used the revised gcForest to perform experiments without data balance.

To compare the performance of weighted-gcForest and gcForest, we set the following parameters on both models. On the scan layer, we used two forests, a random forest and a completely random forest. These two forests each contained 50 trees. On the cascade layer, we used four forests, two random forests and two completely random forests, each of which contained 100 trees. Then, we conducted 10 times of 10-fold cross-validation to evaluate the model, and we calculated the mean and variance as the indicator. We used three indicators: accuracy (ACC), sensitivity (SEN) and specificity (SPE).

B. RESULTS AND ANALYSIS

1) RESULTS OF THE REVISED GCFOREST METHOD

The Cobre dataset was not balanced and could be used to compare the performance of the traditional gcForest with that of the other methods. According to table 1, the accuracy and sensitivity values of the revised gcForest on the Cobre dataset are improved. Compared with the traditional gcForest, the weighted gcForest we propose has a better performance on the Cobre dataset. This indicates that the weighted class vector and weighted prediction vector perform well in the fMRI dataset because of the contribution of different forest representations to the decision vectors.

Figure 7 shows the weight of each forest. Before weight averaging, the weight of each forest is 0.25. Figure 7 shows the difference in weights before and after weight averaging. It is clear that Forest 4 has the highest weight so that its contribution is the highest.

TABLE 1. Comparison results for the revised gcForest and the traditional gcForest for the Cobre dataset without data balancing.

	weighted-gcForest	gcForest
ACC	**0.521±0.360**	0.507±0.312
SEN	**0.640±0.774**	0.607±0.770
SPE	0.520±0.720	**0.563±0.792**

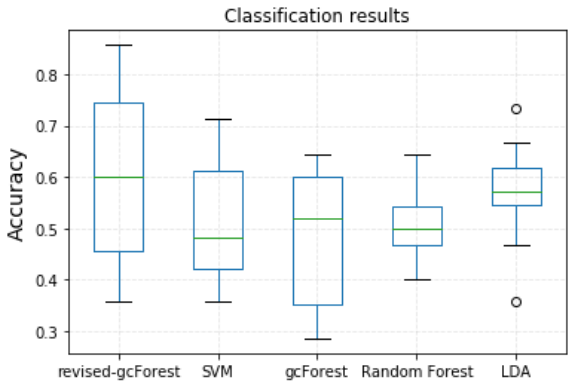


FIGURE 8. The accuracy of each method.

Compared with other methods, Figure 8 shows that the revised method has a better performance on average accuracy. In Figure 8, the revised gcForest has the highest median accuracy and the highest accuracy, but it has the highest variance, which indicates risk. LDA's accuracy is lower than that of the revised gcForest but it has the lowest variance. It is necessary to reduce the variance of the revised gcForest. All in all, the revised gcForest has a better performance on fMRI data.

2) IMPACT OF DATA BALANCING

From table 2, we find the following:

On the unbalanced UCLA data set, data balancing plays an important role. We can see that the average of the UCLA SPE score is 17.0% without data balancing. This is because the model has obvious overfitting, and the test set is fully predicted as a positive example. After the data were balanced, the SPE score on the test set changed from 17.0% to 46.1%, which indicates that the model's overfitting problem was improved.

Comparing the results in the two tables, we find that the revised gcForest is better than the traditional gcForest on all three datasets, and data balancing improves the performance of gcForest on fMRI data. Therefore, the revised gcForest model has a greater impact.

TABLE 2. Comparison results for the revised gcForest with and without data balancing.

	Balanced	Both data	UCLA
ACC	YES	**0.610±0.425**	**0.677±0.428**
	NO	0.601±0.196	0.634±0.414
SEN	YES	0.635±0.622	0.833±0.614
	NO	0.930±0.213	0.893±0.596
SPE	YES	0.545±0.690	0.461±0.838
	NO	0.125±0.514	0.170±0.592

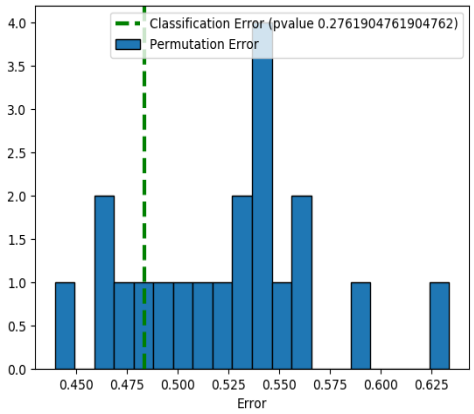


FIGURE 9. The result of permutation test on Cobre dataset.

TABLE 3. Results of replacement tests on different dataset.

Datasets	accuracy	p-value
Both data	0.632	0.276
UCLA data	0.716	0.276
Cobre data	0.510	**0.209**

3) PERMUTATION TEST

Permutation test [26] is used to evaluate the significance of a cross-validated score with permutations. We set r=5 and k=20 in the permutation test to calculate the p value. We randomly shuffle the processed data. The shuffled data is used as the original data. We calculate the 10-fold cross-validation error of the original data r times, and call them 'error original'. We will randomly scramble the original labels k times to calculate k errors. Then we calculate r p-values according to formula 2 in the paper, and take the average value of p-values. This method is used to solve the problem of error instability. We have carried out many experiments. The result is as follows:

In figure 9, the green line represents the original classification error, and the height of the blue block represents the number of errors in the corresponding error rate range.

From the table above, we can see that on the balanced data set Cobre and the unbalanced data set UCLA, the p-value calculated by our model are close, which shows that our method is applicable to both balanced and unbalanced data sets.

IV. CONCLUSION

In this paper, we proposed a revised gcForest to classify patients with schizophrenia and control subjects. To solve the problem of unbalanced data, we randomly cut the training set before SMOTE to reduce the number of noisy points. Experimental results show that our method has superior performance to that of the traditional gcForest on fMRI data. Moreover, we proposed a weight averaging method using the softmax function to average class vectors and prediction class vectors. Experimental results show that it can improve the accuracy of gcForest on fMRI datasets.

In our experiment, the original data are changed. The sample size is changed by data balancing, and the feature size is changed by PCA and multigranularity scanning. These transformations aim to train a better model to promote accuracy.

In the Cobre data set, the numbers of schizophrenia patients and the healthy control group are unbalanced, and we can see from table 2 that the method proposed in this paper has a improvement on this data set. In the UCLA data set, the numbers of schizophrenia patients and the healthy control group are very unbalanced, and we can see from table 2 that the method proposed in this paper has a significant improvement on this data set. Our weighted-gcForest can also be applied to other disease diagnoses with fMRI data, such as ADHD and Alzheimer's disease.

As stated above, the revised gcForest has a high variance. The problem of how to reduce the variance should be taken into account in the future. In this paper, the disadvantage of the softmax function is that it cannot distinguish classifiers with poor generalization ability. But from the experimental results, this method has better results than the traditional gcForest. Furthermore, the error back-propagation (BP) algorithm can be considered in dynamically adjusting the weight of each forest. At last, there may be a better weighting method in the future.

REFERENCES

[1] Y. Mao, "Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3β/β-catenin signaling," *Cell*, vol. 136, no. 6, pp. 1017–1031, 2009, doi: 10.1016/j.cell.2008.12.044.

[2] N. C. Andreasen and W. T. Carpenter, Jr., "Diagnosis and classification of schizophrenia," *Schizophrenia Bull.*, vol. 19, no. 2, pp. 199–214, Jan. 1993, doi: 10.1093/schbul/19.2.199.

[3] N. C. Andreasen, "The diagnosis of schizophrenia," *Schizophrenia Bull.*, vol. 13, no. 1, pp. 9–22, Jan. 1987, doi: 10.1093/schbul/13.1.9.

[4] S. R. Kay, A. Fiszbein, and L. A. Opler, "The positive and negative syndrome scale (PANSS) for schizophrenia," *Schizophrenia Bull.*, vol. 13, no. 2, pp. 261–276, Jan. 1987, doi: 10.1093/schbul/13.2.261.

[5] E. Veronese, U. Castellani, D. Peruzzo, M. Bellani, and P. Brambilla, "Machine learning approaches: From theory to application in schizophrenia," *Comput. Math. Methods Med.*, vol. 2013, pp. 1–12, Nov. 2013, doi: 10.1155/2013/867924.

[6] H. Cheng, S. Newman, J. Goñi, J. S. Kent, J. Howell, A. Bolbecker, A. Puce, B. F. O'Donnell, and W. P. Hetrick, "Nodal centrality of functional network in the differentiation of schizophrenia," *Schizophrenia Res.*, vol. 168, nos. 1–2, pp. 345–352, Oct. 2015, doi: 10.1016/j.schres.2015.08.011.

[7] J. Kim, V. D. Calhoun, E. Shim, and J.-H. Lee, "Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia," *NeuroImage*, vol. 124, pp. 127–146, Jan. 2016, doi: 10.1016/j.neuroimage.2015.05.018.

[8] W. Yan, S. Plis, V. D. Calhoun, S. Liu, R. Jiang, T.-Z. Jiang, and J. Sui, "Discriminating schizophrenia from normal controls using resting state functional network connectivity: A deep neural network and layer-wise relevance propagation method," in *Proc. IEEE 27th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2017, pp. 1–6, doi: 10.1109/mlsp.2017.8168179.

[9] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, "3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI," *IEEE Access*, vol. 5, pp. 23626–23636, 2017, doi: 10.1109/ACCESS.2017.2762703.

[10] F. Zhu, F. Liu, W. Guo, J. Chen, Q. Su, Z. Zhang, H. Li, X. Fan, and J. Zhao, "Disrupted asymmetry of inter- and intra-hemispheric functional connectivity in patients with drug-naive, first-episode schizophrenia and their unaffected siblings," *EBioMedicine*, vol. 36, pp. 429–435, Oct. 2018, doi: 10.1016/j.ebiom.2018.09.012.

[11] M. N. I. Qureshi, J. Oh, and B. Lee, "3D-CNN based discrimination of schizophrenia using resting-state fMRI," *Artif. Intell. Med.*, vol. 98, pp. 10–17, Jul. 2019, doi: 10.1016/j.artmed.2019.06.003.

[12] Z.-H. Zhou and J. Feng, "Deep forest: Towards an alternative to deep neural networks," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3553–3559, doi: 10.24963/ijcai.2017/497.

[13] L. Shao, D. Zhang, H. Du, and D. Fu, "Deep forest in ADHD data classification," *IEEE Access*, vol. 7, pp. 137913–137919, 2019, doi: 10.1109/access.2019.2941515.

[14] K. Sakai and K. Yamada, "Machine learning studies on major brain diseases: 5-year trends of 2014–2018," *Jpn. J. Radiol.*, vol. 37, no. 1, pp. 34–72, 2019, doi: 10.1007/s11604-018-0794-4.

[15] R. A. Poldrack, E. Congdon, W. Triplett, K. J. Gorgolewski, K. H. Karlsgodt, J. A. Mumford, F. W. Sabb, N. B. Freimer, E. D. London, T. D. Cannon, and R. M. Bilder, "A phenome-wide examination of neural and cognitive function," *Sci. Data*, vol. 3, no. 1, Dec. 2016, Art. no. 160110, doi: 10.1038/sdata.2016.110.

[16] K. J. Gorgolewski, J. Durnez, and R. A. Poldrack, "Preprocessed consortium for neuropsychiatric phenomics dataset," *F1000Research*, vol. 6, p. 1262, Sep. 2017, doi: 10.12688/f1000research.11964.2.

[17] C.-G. Yan, X.-D. Wang, X.-N. Zuo, and Y.-F. Zang, "DPABI: Data processing & analysis for (resting-state) brain imaging," *Neuroinformatics*, vol. 14, no. 3, pp. 339–351, Jul. 2016, doi: 10.1007/s12021-016-9299-4.

[18] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, Jan. 2002, doi: 10.1006/nimg.2001.0978.

[19] K. J. Friston, "Functional and effective connectivity: A review," *Brain Connectivity*, vol. 1, no. 1, pp. 13–36, Jan. 2011, doi: 10.1089/brain.2011.0008.

[20] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Mach. Learn.*, vol. 40, no. 2, pp. 139–157, Aug. 2000, doi: 10.1023/a:1007607513941.

[21] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901, doi: 10.1080/14786440109462720.

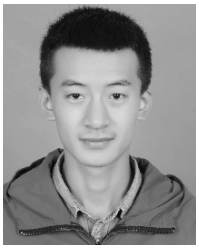
[22] R. Guo, S.-Y. Xie, X.-N. Cheng, and H. Zhao, "Combined SVM and PCA to recognize the brain function from fMRI images," in *Proc. 3rd Int. Conf. Bioinf. Biomed. Eng.*, Jun. 2009, pp. 1–3, doi: 10.1109/ICBBE.2009.5162526.

[23] S.-Y. Xie, R. Guo, N.-F. Li, G. Wang, and H.-T. Zhao, "Brain fMRI processing and classification based on combination of PCA and SVM," in *Proc. Int. Joint Conf. Neural Netw.*, Jun. 2009, pp. 3510–3515, doi: 10.1109/IJCNN.2009.5179085.

[24] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newslett.*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.

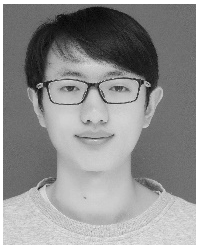
[25] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[26] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance," *J. Mach. Learn. Res.*, vol. 11, pp. 1833–1863, Jun. 2010, doi: 10.1109/ICDM.2009.108.



YAFEI ZHU is currently pursuing the B.S. degree with the School of Computer Science and Technology, Southwest Minzu University, China. His current research interests include machine learning and medical imaging.

SHUYUE FU is currently pursuing the B.S. degree with the School of Computer Science and Technology, Southwest Minzu University, China. Her current research interests include artificial intelligence and medical imaging.



SHIHU YANG is currently pursuing the B.S. degree with the School of Computer Science and Technology, Southwest Minzu University, China. His current research interests include machine learning and artificial intelligence.

PING LIANG received the M.Sc. degree in software engineering from the University of York, U.K., in 2002, and the Ph.D. degree in computer science from Khon Kaen University, Thailand, in 2017. She is currently a Lecturer with Southwest Minzu University, Chengdu, China. Her current research interests include machine learning and artificial intelligence.

YING TAN received the M.S. and Ph.D. degrees in mechanical engineering from Sichuan University, in 2002 and 2008, respectively. He is currently a Professor with Southwest Minzu University, Chengdu, China, where he is the Director of the Key Laboratory for Computer Systems of State Ethnic Affairs Commission. His research interests include artificial intelligence and functional neuroimaging.