Supplementary material

EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Conditions

Datasets

In this paper, Sakar [19], MaxLittle [23] and the self-collected PD speech dataset (named SelfData) [5] are used as the target datasets for the validation of the proposed algorithm.

These three datasets are highly representative, Sakar and MaxLittle are mainly used for diagnosis, and SelfData is mainly used for efficacy assessment. Sakar and MaxLittle are public PD speech datasets based on European and American patients, and SelfData is a self-collected PD speech dataset based on Chinese patients. Thus, these three datasets cover a representative range of patients from distinct applications and different regions. Table I shows the basic information of the three datasets.

Sakar dataset ('Parkinson Speech Dataset with Multiple Types of Sound Recordings Data Set', for short, called 'Sakar' dataset) was created by Sakar et al. This dataset contained a total of 40 subjects, including 20 patients with Parkinson's disease (6 females and 14 males) and 20 healthy subjects (10 females and 10 males). Each subject was provided with 26 corpus samples. Twenty-six features were extracted from each corpus sample as a feature vector. For more detailed information, please refer to [19].

MaxLittle dataset ('Parkinson Telemonitoring Data Set', for short, called 'MaxLittle') was created by MaxLittle et al. This dataset contained a total of 31 subjects, including 23 patients with Parkinson's disease (7 females and 16 males) and 8 healthy subjects (5 females and 3 males). Each subject was provided with 6 or 7 corpus samples. Twenty-two features were extracted from each corpus sample as a feature vector. For more detailed information, please refer to [23].

SelfData dataset was provided by the First Affiliated Hospital (Southwest hospital) of the Army Medical University, Chongqing, China. This dataset contained recordings of 36 PD patients (16 female and 20 male) without receiving treatment and 54 PD patients (27 female and 27 male) after receiving medication. Each subject was provided with 13 corpus samples. Twenty-six features were extracted from each corpus sample as a feature vector. For more detailed information, please refer to [5].

In the SelfData dataset, the age distribution of male patients ranged from 39 to 82 years old and that of female patients ranged from 45 to 76 years old, which means that the age range of Parkinson's disease patients is extremely widely distributed. Therefore, proposing an algorithm that can be efficiently and conveniently applied to the early diagnosis of Parkinson's disease has practical application and helps to promote the development of Parkinson's disease diagnosis and treatment.

Experimental Configuration

The SVM, KNN and ELM classifiers were used in the experiments. For fair comparison, default settings are used for all classifier parameters in this paper, so that different algorithms based on the same classifier can be effectively compared to verify the effectiveness of the proposed algorithm.

To reflect the accuracy of the algorithm more comprehensively, two cross-validation methods, Leave One Subject Out (LOSO) and Holdout, were used for experimental validation. LOSO, Holdout and K-fold all belong to cross-validation methods. Among them, LOSO and Holdout are more suitable for datasets with fewer samples. In the case of fewer samples, its accuracy rate is closer to the actual clinical situation. Most of the compared methods also use these two cross-validation methods, so this paper chooses these two cross-validation methods for fair comparison. In this paper, we evaluate the performance of each method based on Accuracy (ACC), Sensitivity (SEN), and Specificity (SPE) criteria. Equipment hardware and software environment: CPU (Intel i3-7100@3.90 GHz), memory 8 GB, MATLAB R2018b.

All the experimental results are shown in the supplementary material which can be found in GitHub (https://github.com/wywwwww/EMSFE-supplementary-material.git).

B. Analysis of The Extracted Features on Envelope Samples

To verify the contribution of the proposed EMSFE to the mentioned above clinical problems, we organized feature extraction experiments and analyzed the extracted features. The three datasets (Sakar, MaxLittle, and SelfData) are involved here. Fig. 4-6 show the feature weight maps of each envelope before and after sample compression, and the weight values of different features in these samples are characterized by different color blocks.

Fig. 4-6(a) show the feature weights of the original intra-subject corpus samples by SS/FSM calculation. It can be clearly seen that the sensitive features are not consistent among the different corpus samples and are sparsely distributed, so the sensitive features of a single corpus sample cannot be used as diagnostic markers for the whole subject. Therefore, we have to compress the corpus samples to find uniform and stable sensitive features that reflect the overall pathology of the subject based on more compact samples.

Fig. 4-6(b) show the feature weights of the compressed samples (envelope samples) processed by EMSFE. Fig. 4-6(c) show the uniform diagnostic markers that reflect the pathology of the whole subject, extracted according to the feature weight values of different samples in Fig. 4-6(b). As seen from Fig. 4-6(b), the number of samples within the envelope is significantly reduced, and the distribution of sensitive features of different samples is more concentrated,

TABLE I
DATASETS INFORMATION

Dataset	Number of subjects	Patients	Healthy subjects	Number of samples per subject	Samples	Features	References
Sakar	40	20	20	26	1040	26	[19]
MaxLittle	31	23	8	6 or 7	189	22	[23]
SelfData	90	36	54	13	1170	26	[5]

which facilitates the extraction of uniform and stable diagnostic features of the subject. For the Sakar dataset, after sample compression and simultaneous sample and feature selection, six high-quality envelope samples were obtained for each subject compared to the original 26 samples. As shown in Fig. 4(c), features 10, 24-26 can effectively reflect the pathology of the entire subject and can therefore be referred to as potential diagnostic markers. Similarly, in the MaxLittle dataset, two high-quality envelope samples were obtained for each subject, and features 3, 13, 19, and 22 can be used as potential diagnostic markers as Fig. 5(c) shows. It can be seen form Fig. 6(c), in the SelfData dataset, four high-quality envelope samples were obtained for each subject, and features 2, 22, 23, 24, and 26 can be used as potential diagnostic markers of the subject. The results show that the envelope samples are helpful to find the sensitive features for whole subject, and the relationship between features and samples become clearer. The proposed method helps to advance the speech diagnosis technique significantly to clinical applications.

C. Ablation Experiments

Verification of The Proposed Algorithm (EMSFE)

In order to verify the generalization performance of the proposed algorithm, 'SVM' and 'EMSFE+SVM' are compared based on the generic datasets and the PD speech datasets. 'SVM' means SVM classification without feature and sample learning. 'EMSFE+SVM' means SVM classification with the proposed algorithm. The generic datasets are frequently used ones and the specific results are shown in Tables II-IV.

TABLE II
VERIFICATION OF EMSFE IN THE GENERIC DATASETS

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Yeast-1-2-8-9-vs. -7	SVM	73.75	67.78	73.95	69.94
	EMSFE +SVM	**74.56**	**71.11**	**74.67**	**72.40**
Yeast-0-5-6-7-9-vs.-4	SVM	83.62	73.57	84.72	77.91
	EMSFE +SVM	**86.63**	**75.00**	**87.92**	**80.18**
Wine	SVM	94.73	91.36	96.97	94.00
	EMSFE +SVM	**95.64**	**93.18**	**97.27**	**95.13**
WDBC	SVM	94.83	90.47	97.41	93.72
	EMSFE +SVM	**95.29**	**91.56**	**97.50**	**94.36**
Glass6	SVM	95.08	**91.11**	95.70	93.25
	EMSFE +SVM	**96.31**	**91.11**	**97.14**	**93.94**

TABLE III
VERIFICATION OF EMSFE IN THE PD SPEECH DATASETS (LOSO)

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar	SVM	52.79	51.73	53.85	52.28
	EMSFE +SVM	**90.00**	**95.00**	**85.00**	**90.48**
MaxLittle	SVM	61.46	70.83	52.08	64.76
	EMSFE +SVM	**93.75**	**100.00**	**87.50**	**94.12**
SelfData	SVM	51.82	47.01	56.62	49.38
	EMSFE +SVM	**87.50**	**88.89**	**86.11**	**87.67**

TABLE IV
VERIFICATION OF EMSFE IN THE PD SPEECH DATASETS (HOLDOUT)

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar	SVM	64.81	64.10	65.51	64.56
	EMSFE +SVM	**94.17**	**96.67**	**91.67**	**94.31**
MaxLittle	SVM	73.53	62.86	84.29	70.40
	EMSFE +SVM	**97.50**	**100.00**	**95.00**	**97.56**
SelfData	SVM	56.14	56.36	55.93	56.24
	EMSFE +SVM	**91.43**	**93.00**	**90.00**	**91.63**

As shown in Tables II-IV, the proposed algorithm has a certain improvement on each evaluation metric in both the generic and PD speech datasets, which means that the proposed algorithm has robust generalization performance. Therefore, the algorithm can not only significantly improve the accuracy of PD speech diagnosis but also can be applied into other fields.

Verification of The Sample Pruning Module (SPM)

To verify the effectiveness of the proposed sample pruning module (SPM), we compared 'SVM' and 'SPM+SVM' based on three PD speech datasets. 'SVM' means SVM classification without feature and sample learning. 'SPM+SVM' means SVM only with a sample pruning process of original dataset. Tables V-VI show the specific results.

TABLE V
VERIFICATION OF THE SAMPLE PRUNING MODULE (LOSO)

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar	SVM	52.79	51.73	53.85	52.28
	SPM +SVM	**53.80**	**53.91**	53.70	**53.85**
MaxLittle	SVM	61.46	70.83	52.08	64.76
	SPM +SVM	**65.00**	**77.50**	**52.50**	**68.89**
SelfData	SVM	51.82	47.01	56.62	49.38
	SPM +SVM	**57.07**	**53.28**	**60.86**	**55.38**

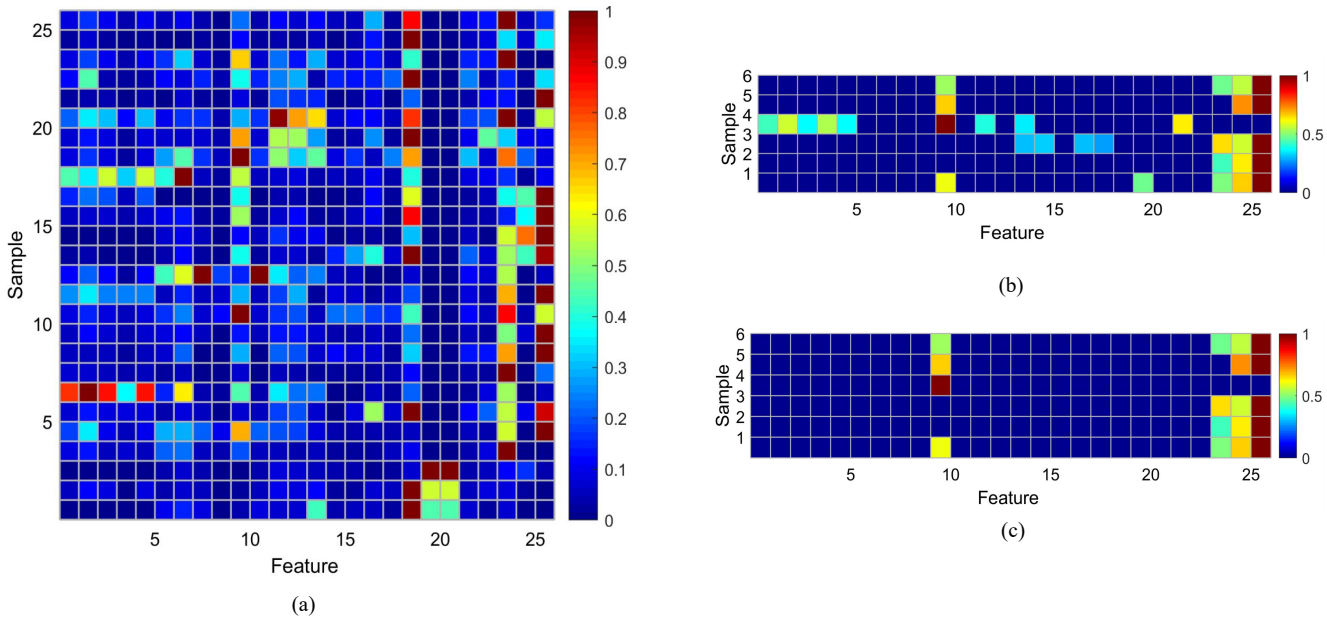

Fig. 4. Feature weight maps (Sakar): (a) feature weight map of the original envelope; (b) feature weight map of the compressed envelope; (c) feature extraction (diagnostic markers) map

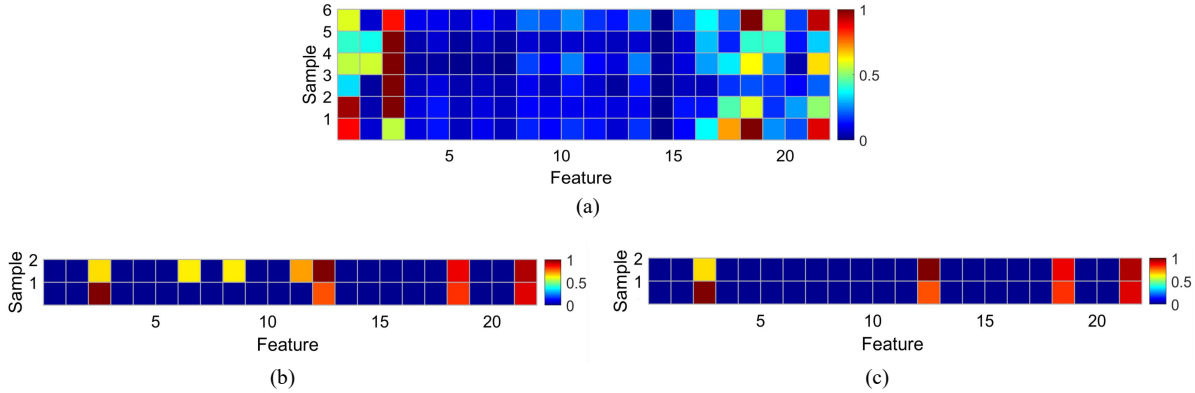

Fig. 5. Feature weight maps (MaxLittle): (a) feature weight map of the original envelope; (b) feature weight map of the compressed envelope; (c) feature extraction (diagnostic markers) map

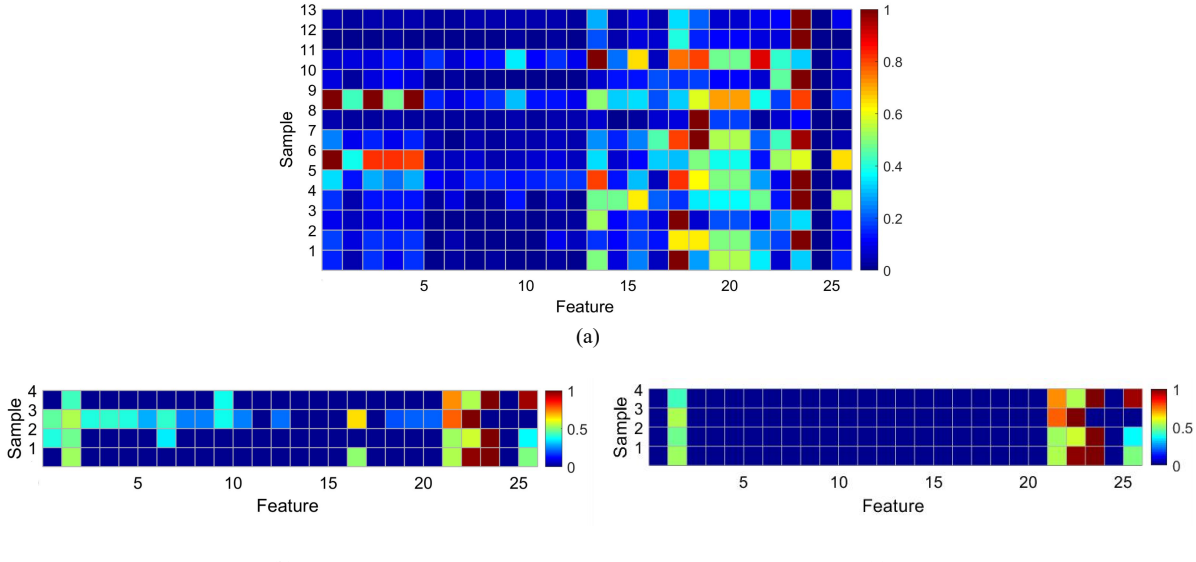

Fig. 6. Feature weight maps (SelfData): (a) feature weight map of the original envelope; (b) feature weight map of the compressed envelope; (c) feature extraction (diagnostic markers) map

TABLE VI
VERIFICATION OF THE SAMPLE PRUNING MODULE (HOLDOUT)

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar	SVM	64.81	64.10	65.51	**64.56**
	SPM +SVM	**65.22**	**65.22**	**65.22**	65.22
MaxLittle	SVM	73.53	62.86	**84.29**	70.40
	SPM +SVM	**74.58**	**70.00**	79.17	**73.36**
SelfData	SVM	56.14	56.36	55.93	56.24
	SPM +SVM	**59.70**	**57.73**	**61.19**	**58.75**

From Tables V-VI, it can be seen that the sample pruning module can generally improve the classification performance in terms of the four evaluation metrics, which means SPM is an effective method for compressing samples.

Verification of The Two-step Sample Compression Module (TSCM)

To verify the effectiveness of the two-step sample compression module (TSCM), we compared 'SVM' and 'TSCM +SVM' based on three PD speech datasets. 'SVM' means SVM classification without feature and sample learning. 'TSCM +SVM' means SVM only with a sample compression process by TSCM. Tables VII-VIII show the results.

TABLE VII
VERIFICATION OF THE TWO-STEP SAMPLE COMPRESSION MODULE (LOSO)

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar	SVM	52.79	51.73	**53.85**	52.28
	TSCM +SVM	**53.52**	**58.81**	48.86	**56.02**
MaxLittle	SVM	61.46	70.83	52.08	64.76
	TSCM +SVM	**85.94**	**96.88**	**75.00**	**87.33**
SelfData	SVM	51.82	**47.01**	56.62	**49.38**
	TSCM +SVM	**53.61**	44.44	**62.78**	48.93

TABLE VIII
VERIFICATION OF THE TWO-STEP SAMPLE COMPRESSION MODULE (HOLDOUT)

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar	SVM	64.81	64.10	65.51	64.56
	TSCM +SVM	**82.50**	**76.67**	**81.16**	**78.43**
MaxLittle	SVM	73.53	62.86	84.29	70.40
	TSCM +SVM	**94.74**	**95.56**	**87.00**	**91.64**
SelfData	SVM	56.14	56.36	55.93	56.24
	TSCM +SVM	**74.31**	**67.78**	**80.83**	**72.51**

As shown in Tables VII-VIII, based on different datasets, the TSCM can generally improve the classification accuracy, and there is also a certain degree of improvement in the SEN, SPE and F1 metrics. Therefore, the proposed TSCM can not only compress the samples, but also effectively improve the classification accuracy.

Verification of Multilayer TSCM (MSCM)

In this paper, we propose a multilayer sample compression module (MSCM) for sample compression based on TSCM for the case where the subjects contain a large number of corpus samples. By performing feature selection and classifier training on the envelope set of each layer obtained after MSCM compression, multiple classifiers can be obtained, and these classifiers are integrated to obtain the final ensemble learning model. To verify the effectiveness of the MSCM, we test the classifiers at each compressed layer based on the LOSO cross-validation method. The specific results are shown in Table IX. Also, 'SS/FSM' and EMSFE are compared based on different classifiers, where 'SS/FSM' means that the original data set is handled by SS/FSM only. The specific results are shown in Tables X-XI.

Table IX shows the true labels of some test subjects, the integrated predicted labels, and the test labels for each layer of the classifier, from these results it can be seen that the proposed algorithm has error correction capability. For example, as shown by subjects 23, 31, and 39, the layer 1 prediction labels are inconsistent with the actual labels, but more than 80% of the layers from layer 2 to layer 6 are correctly predicted and can be finally corrected by integration. However, there are also errors, as shown in subject 8, where the layer 1 prediction label are consistent with the actual label, but the final prediction label obtained by SFELM is inconsistent with the actual label, resulting in misclassification. Overall, the number of successful error corrections is much greater than the number of misclassifications apparently. Therefore, it can be seen that the MSCM has error correction capability and can improve the classification accuracy.

As shown in Tables X-XI, the performance of the EMSFE is generally significantly better than that of the 'SS/FSM' algorithm in terms of four evaluation metrics. For example, 'EMSFE+SVM' obtains the better performance on Sakar with the mean ACC, SEN, SPE and F1 results of 90%, 95%, 85%, and 90.48%, which are all around 10% better than 'SS/FSM+SVM'. It means that the proposed algorithm not only has a certain error correction capability but also can improve classification accuracy by integrating classifiers trained on envelope sets with different compression levels.

D. Algorithm Comparison

Comparison with Classical Algorithms

The proposed algorithm (EMSFE) contains the ideas of integration and sample transformation, so the effectiveness of this algorithm is verified by comparing it with representative ensemble learning algorithm (XGBoost) and sample transformation algorithm (PD speech transformation algorithm) respectively. The specific results are shown in Tables XII-XIII.

(1) Comparison with XGBoost

XGBoost is an important ensemble learning algorithm. The proposed algorithm generates the envelope sets with different compression levels by MSCM, and integrates the classifiers trained on the envelope sets at each layer by SFELM, so as to improve the classification performance of the model. The experimental results of the comparison of the two above algorithms are presented in Table XII.

TABLE IX
ERROR CORRECTION PERFORMANCE OF EMSFE (SAKAR)

Subject num	Actual label	Predict label	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	1	1	1	0	0	1	1
4	0	0	0	0	1	0	0	1
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**
9	0	0	0	0	1	0	0	0
10	0	1	1	1	1	1	1	1
11	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0
13	0	0	0	0	1	0	0	0
14	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0
16	0	0	0	1	0	1	1	0
17	0	0	0	0	0	0	0	0
18	0	0	0	0	0	1	1	0
19	0	0	0	0	0	0	0	0
20	0	0	0	1	0	0	1	0
21	1	1	1	1	1	1	1	1
22	1	1	1	1	1	0	1	1
23	**1**	**1**	**0**	**1**	**1**	**0**	**1**	**1**
24	1	1	1	1	1	0	0	1
25	1	1	1	1	1	1	1	1
26	1	1	1	1	1	1	1	1
27	1	1	1	0	0	1	0	1
28	1	1	1	1	0	1	0	1
29	1	1	1	1	0	0	0	0
30	1	1	1	1	1	1	1	1
31	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**
32	1	1	1	0	1	1	1	1
33	1	1	1	1	1	1	1	1
34	1	1	1	1	1	0	1	1
35	1	1	1	0	0	0	1	1
36	1	1	1	1	1	1	1	1
37	1	1	1	1	1	1	1	1
38	1	1	1	1	1	1	0	1
39	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**
40	1	0	0	1	0	0	0	0

Note: red: successful error correction examples; yellow: misclassification examples

TABLE X
VERIFICATION OF THE MULTILAYER SAMPLE COMPRESSION MODULE (LOSO)

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar	SS/FSM +SVM	80.00	85.00	75.00	80.95
	EMSFE +SVM	**90.00**	**95.00**	**85.00**	**90.48**
	SS/FSM +KNN	82.50	80.00	85.00	82.05
	EMSFE +KNN	**90.00**	**90.00**	**90.00**	**90.00**
	SS/FSM +ELM	85.00	85.00	85.00	85.00
	EMSFE +ELM	**92.50**	**90.00**	**95.00**	**92.31**
MaxLittle	SS/FSM +SVM	**93.75**	**100.00**	**87.50**	**94.12**
	EMSFE +SVM	**93.75**	**100.00**	**87.50**	**94.12**
	SS/FSM +KNN	81.25	87.50	75.00	82.35
	EMSFE +KNN	**93.75**	**100.00**	**87.50**	**94.12**
	SS/FSM +ELM	93.75	87.50	**100.00**	93.33
	EMSFE +ELM	**96.87**	**93.75**	**100.00**	**96.77**
SelfData	SS/FSM +SVM	77.78	77.78	77.78	77.78
	EMSFE +SVM	**87.50**	**88.89**	**86.11**	**87.67**
	SS/FSM +KNN	72.22	63.89	80.56	69.70
	EMSFE +KNN	**90.28**	**83.33**	**97.22**	**89.55**
	SS/FSM +ELM	73.61	75.00	72.22	73.97
	EMSFE +ELM	**91.67**	**88.89**	**94.44**	**91.43**

TABLE XI
VERIFICATION OF THE MULTILAYER SAMPLE COMPRESSION MODULE (HOLDOUT)

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar	SS/FSM +SVM	87.50	93.33	81.67	88.19
	EMSFE +SVM	**94.17**	**96.67**	**91.67**	**94.31**
	SS/FSM +KNN	90.83	88.33	**93.33**	90.60
	EMSFE +KNN	**93.33**	**93.33**	**93.33**	**93.33**
	SS/FSM +ELM	95.83	**96.67**	95.00	95.87
	EMSFE +ELM	**98.33**	**96.67**	**100.00**	**98.30**
MaxLittle	SS/FSM +SVM	93.75	92.50	**95.00**	93.67
	EMSFE +SVM	**97.50**	**100.00**	**95.00**	**97.56**
	SS/FSM +KNN	95.00	95.00	95.00	95.00
	EMSFE +KNN	**97.50**	**97.50**	**97.50**	**97.50**
	SS/FSM +ELM	**98.75**	97.50	**100.00**	**98.73**
	EMSFE +ELM	97.50	**100.00**	95.00	97.56
SelfData	SS/FSM +SVM	82.86	79.09	87.00	82.35
	EMSFE +SVM	**91.43**	**93.00**	**90.00**	**91.63**
	SS/FSM +KNN	84.76	**93.67**	75.00	85.67
	EMSFE +KNN	**94.29**	92.00	**96.36**	**94.05**
	SS/FSM +ELM	81.90	77.27	87.00	81.22
	EMSFE +ELM	**94.29**	**98.00**	**90.91**	**94.64**

TABLE XII
COMPARISON WITH XGBOOST (HOLDOUT)

Dataset	Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar	XGBoost	68.91	66.03	71.79	67.99
	EMSFE	**94.17**	**96.67**	**91.67**	**94.31**
MaxLittle	XGBoost	93.22	97.78	78.57	89.21
	EMSFE	**96.87**	**93.75**	**100.00**	**96.77**
SelfData	XGBoost	62.10	42.55	75.24	50.86
	EMSFE	**91.43**	**93.00**	**90.00**	**91.63**

It can be clearly seen from Table XII that the proposed algorithm obtains the better performance on different data sets and evaluation metrics. This means that the proposed algorithm is able to obtain envelope sets containing sample information of different compression levels, thus increasing the diversity of classifiers and greatly improving model performance through integration.

(2) Comparison with the PD speech transformation algorithm

Sakar [19] used a simple linear sample transformation method that utilized concentrated trends and discrete measures such as the mean, median, trimmed mean (removing 10% and 25%), standard deviation, interquartile range, and mean absolute deviation for 26 speech samples from each subject. The proposed MSCM compresses corpus samples based on TSCM with sample selection followed by sample transformation, thereby generating envelope sets with different compression levels. The samples and features are selected simultaneously for envelope set of each layer, thus obtaining a small number of high-quality samples and features that can characterize the pathology of the subjects. The experimental results of the comparison of the above two PD speech sample transformation algorithms are shown in Table XIII.

TABLE XIII
COMPARISON WITH THE PD SPEECH SAMPLE TRANSFORMATION ALGORITHM (SAKAR)

Method	ACC (%)	SEN (%)	SPE (%)	F1 (%)
Sakar et al. [19] +KNN +LOSO	54.04	53.27	54.81	53.68
EMSFE +KNN +LOSO	**90.00**	**90.00**	**90.00**	**90.00**
Sakar et al. [19] +SVM +LOSO	52.50	52.50	52.50	52.50
EMSFE +SVM +LOSO	**90.00**	**95.00**	**85.00**	**90.48**

As shown in Table XIII, the proposed algorithm is effective and has a large improvement in four evaluation metrics compared to the simple linear sample transformation. This means that the proposed algorithm can obtain a small number of high-quality samples through sample selection and sample transformation, thus improving the performance of the overall model.

Comparison with State-of-the-art Algorithms

To further verify the performance of the proposed EMSFE, we compare the proposed algorithm with state-of-the-art algorithms in the representative literature on PD diagnosis and the results are shown in Tables XIV-XVI. For fair comparison, Table XIV presents the results of the proposed algorithm for the Sakar dataset using three cross-validation methods-LOSO, Holdout, and 10-fold CV. Table XV presents the results of the proposed algorithm for the MaxLittle dataset using the Holdout cross-validation method. Table XVI presents the results of the proposed algorithm for the SelfData dataset using the LOSO and Holdout cross-validation.

It can be seen that the proposed algorithm still has shown improvement in the evaluation metrics, compared with the existing PD diagnosis advanced literature algorithms. It means that the proposed algorithm can improve the model classification performance by sample compression and integrating envelope sets with different compression levels.

E. Effect Analysis of Layers of Multilayer Sample Compression Module (MSCM)

The number of TSCM is the number of layer L . To investigate the effect of the number of TSCM on the performance of EMSFE, three datasets, Sakar, MaxLittle and SelfData, are selected for parametric analysis. Fig. 7 show the ACC, SEN, SPE and Matthews correlation coefficient (MCC) of the proposed algorithm based on LOSO cross-validation method with different deep sample fusion layers. The number of fusion layers is from the original to the maximum fusion layer. In addition, the datasets used in this paper, MaxLittle and SelfData, are two imbalanced datasets. For the imbalanced datasets, we use the under-sampling method to balance the data, and the classification performance with and without balance processing are shown in Fig. 7.

As shown in Fig. 7, the classification performance of the three datasets tends to increase as the number of layers of deep sample space fusion increases, which indicates that the classifier trained by fusing subsets of deep sample space based on different compression levels can effectively improve the overall model performance. For the MaxLittle and SelfData datasets, the datasets with the balancing process outperformed the classification performance of the datasets without the balancing process in terms of each evaluation metric. Therefore, the balanced MaxLittle and SelfData datasets are used in the subsequent comparison experiments in this paper. The number of deep sample space fusion layers for the Sakar, balanced MaxLittle, and balanced SelfData datasets were set to 6, 3, and 4, respectively, by considering the evaluation metrics, model complexity, and computation time cost.

TABLE XIV
COMPARISON WITH ADVANCED PD DIAGNOSIS ALGORITHMS (SAKAR)

Study	Method	ACC (%)	SEN (%)	SPE (%)
Sakar et al. [19]	KNN+SVM	55.00 (LOSO)	60.00	50.00
Canturk and Karabiber [26]	4 Feature Selection Methods+6 Classifiers	57.50 (10-fold CV)	54.28	80.00
Zhang et al. [27]	MENN+RF with MENN	81.50 (LOSO)	92.50	70.50
Benba et al. [28]	HFCC+SVM	87.50 (LOSO)	90.00	85.00
Li et al. [20]	Hybrid feature learning+SVM	82.50 (LOSO)	85.00	80.00
Benba et al. [29]	MFCC+SVM	82.50 (LOSO)	80.00	85.00
Behroozi and Sami [30]	Multiple classifier framework	87.50 (LOSO)	90.00	85.00
Zhang [31]	LSVM+MSVM+RSVM+ CART+KNN+LDA+NB	94.17 (Holdout)	50.00	94.92
Khan et al. [32]	Evolutionary neural network ensembles	90.00 (10-fold CV)	93.00	97.00
Soumaya et al. [33]	GA+SVM	91.18 (10-fold CV)	—	—
Proposed algorithm	EMSFE+SVM	**90.00 (LOSO)** **94.17 (Holdout)** **100.00 (10-fold CV)**	**95.00** **96.67** **100.00**	**85.00** 91.67 **100.00**

TABLE XV
COMPARISON WITH ADVANCED PD DIAGNOSIS ALGORITHMS (MAXLITTLE)

Study	Method	ACC (%)	SEN (%)	SPE (%)
Luukka. [34]	Fuzzy entropy measures+similarity	85.03 (Holdout)	—	—
Spadoto et al. [35]	PSO+OPF harmony search+OPF gravitational search+OPF	84.01 (Holdout)	—	—
Das [36]	ANN decision tree	92.90 (Holdout)	—	—
Daliri [37]	SVM with Chi-square distance	91.20 (Holdout)	91.71	89.92
Al-Fatlawi et al. [11]	DBN	91.25 (Holdout)	90.50	93.00
Kadam and Jadhav [38]	FESA-DNN	93.84 (Holdout)	95.23	90.00
Senturk [39]	FS+SVM	93.84(Holdout)	—	—
Sharma et al. [40]	MGWO	93.87(Holdout)	—	—
Despotovic et al. [41]	GPC-ARD exponential	93.29(10-fold CV)	84.00	96.52
Lamba et al. [42]	Genetic algorithm	95.58(10-fold CV)	93.19	97.95
Proposed algorithm	EMSFE+SVM	**97.50 (Holdout)**	**100.00**	95.00

TABLE XVI
COMPARISON WITH ADVANCED PD DIAGNOSIS ALGORITHMS (SELFDATA)

Study	Method	ACC (%)	SEN (%)	SPE (%)
Ali et al. [10]	LDA-NN-GA	63.00 (LOSO)	—	—
Liu et al. [5]	LPP+RF	56.33 (LOSO)	45.58	29.17
Yang et al. [43]	KPCA-SVM	59.67 (LOSO)	—	—
Galaz et al. [44]	SFFS-RF	60.00 (LOSO)	—	—
Cigdem and Demirel [45]	ReliefF-FC-SVM	62.67 (LOSO)	—	—
Proposed algorithm	EMSFE+SVM	**87.50 (LOSO)** **91.43 (Holdout)**	**88.89** **93.00**	**86.11** **90.00**

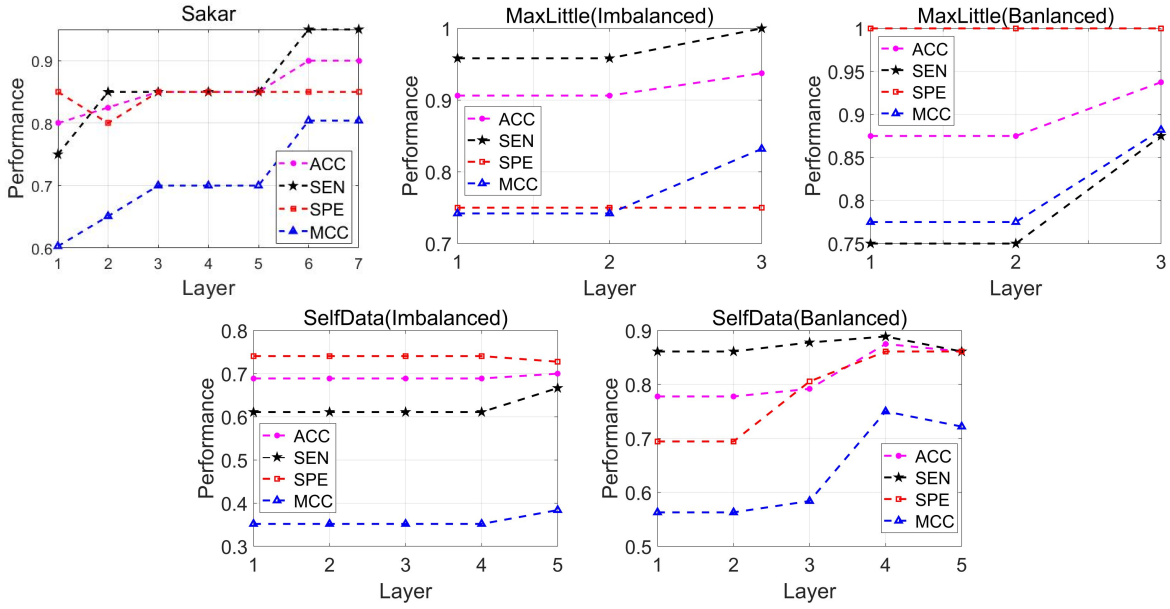

Fig. 7. EMSFE performance with different L