Supplement Material

Chuanyan Zhou1, Jie Ma1, Fan Li, Yongming Li1\*, *Member*, *IEEE*, Pin Wang1

# IV Proposed Method

## Division of training and test set

After dividing the training data according to the method described in this paper, the performance of ES-ESAE is verified by test data. First, the test data  are input into the first trained SPC to obtain . Then, in PS,  is input into the first layer of trained ICMC to obtain , where . In SS,  is input into the second layer of the trained ICMC to obtain , where . Then,  and  are input into the second and third trained SPC to obtain  and  in the PS and SS, respectively. Finally, three layers of envelope samples (,and) will be input into each layer of the trained ESAE to extract deep mixed features F1, F2 and F3 and to obtain classification results. The results from three layers of envelope samples (,and) are fused to obtain the final results.

# V Experimental results and analysis

This section presents the experimental results for the additional datasets. The data and codes can be found in <https://github.com/ChuanyanZhou/NEESAE>.

## Experimental conditions

In this paper, the performance of the proposed algorithm was tested on 15 representative data sets, among which 9 data sets are presented in the article, and the rest 6 data sets are shown in the Table I.

TABLE I

Basic information of the additional data sets used in the study

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | Instances | Attributes | Class |
| Pen-Based Recognition of Handwritten Digit (Pendigits) | 10992 | 16 | 10 |
| Statlog Vehicle Silhouettes (Vehicle) | 846 | 18 | 4 |
| Statlog Heart Dataset (Heart) | 270 | 13 | 2 |
| Maxlettle Parkinson Dataset (Maxlettle) | 195 | 22 | 2 |
| Breast Cancer Wisconsin Original (Wisconsin) | 683 | 9 | 2 |
| Pima Indians Diabetes Dataset (PID) | 768 | 8 | 2 |

## Ablation study

*1) Effectiveness analysis of SPC*

TABLE II

Comparison of the original sample and sample-pair

(on the additional data sets)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset |  | OF (%) | OF&PCA (%) | OF&LDA (%) |
| Pendigits | Original sample | **98.13±0.05** | **98.07±0.13** | **97.87±0.23** |
| Sample-pair | 66.36±0.98 | 68.56±2.40 | 64.63±1.84 |
| Vehicle | Original sample | 80.35±1.31 | 82.34±1.05 | 82.55±0.92 |
| Sample-pair | **83.90±0.19** | **85.32±0.78** | **85.53±1.05** |
| heart | Original sample | 80.89±4.26 | 85.11±3.30 | 83.33±3.60 |
| Sample-pair | **85.56±2.83** | **90.67±2.02** | **90.67±2.30** |
| Maxlittle | Original sample | 85.54±4.01 | 88.00±2.28 | 88.62±4.16 |
| Sample-pair | **86.77±2.57** | **91.08±3.15** | **90.78±1.09** |
| Wisconsin | Original sample | 96.30±1.72 | 97.18±1.19 | 96.83±1.26 |
| Sample-pair | **97.18±1.48** | **98.06±0.74** | **97.89±1.26** |
| PID | Original sample | 70.39±2.74 | 72.34±1.98 | 75.78±3.49 |
| Sample-pair | **74.14±4.27** | **80.16±1.78** | **82.34±1.88** |

*2) Effectiveness analysis of ICMC*

TABLE III

Comparison of sample-pair and sample-pair &ICMC

(on the additional data sets)

|  |  |  |
| --- | --- | --- |
| Dataset | Sample-pair (%) | Sample-pair &ICMC (%) |
| Pendigits | 66.36±0.98 | **69.61±1.62** |
| Vehicle | **83.90±0.19** | 77.66±0.43 |
| heart | 85.56±2.83 | **92.00±3.08** |
| Maxlittle | 86.77±2.57 | **87.69±3.61** |
| Wisconsin | 97.18±1.48 | **98.24±0.82** |
| PID | 74.14±4.27 | **79.14±1.42** |

*3) Effectiveness analysis of the NSELM and ESAE*

TABLE IV

NSELM and ESAE effectiveness analysis experimental comparison

(on the additional data sets)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | NE\_ESAE without NSELM (%) | NE\_ESAE  without  ESAE (%) | | NE\_ESAE (%) | |
| Pendigits | **98.51±0.27** | | 67.40±1.49 | | 84.21±0.83 | |
| Vehicle | 78.58±6.70 | 80.99±2.85 | | **89.53±1.96** | |
| Heart | 84.44±5.88 | 96.89±2.53 | | **99.44±1.11** | |
| Maxlittle | 86.15±3.08 | 89.23±5.76 | | **96.92±1.09** | |
| Wisconsin | 97.35±0.88 | 99.50±0.48 | | **99.89±0.21** | |
| PID | 73.44±7.14 | 82.50±2.28 | | **89.38±1.88** | |

*4) Effectiveness analysis of MSEM*

TABLE IV

MSEM effectiveness analysis experimental comparison

(on the additional data sets)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | OS (%) | PS (%) | SS (%) | MSEM (%) | |
| MV | WF |
| Pendigits | 83.74±0.62 | 67.47±5.11 | 70.93±6.92 | 79.61±7.99 | **84.21±0.83** |
| Vehicle | 80.08±2.49 | 83.59±4.96 | 86.02±3.79 | 87.66±2.97 | **89.53±1.96** |
| heart | 92.52±3.19 | 94.72±2.46 | 99.17±1.06 | 97.78±2.57 | **99.44±1.11** |
| Maxlittle | 95.08±2.28 | 91.38±6.21 | 94.78±4.56 | **96.92±1.09** | 96.25±5.07 |
| Wisconsin | 99.19±1.28 | 98.68±0.88 | 99.11±0.87 | **99.89±0.21** | 98.88±3.07 |
| PID | 80.08±2.49 | 85.31±1.58 | 86.41±3.80 | 88.20±2.63 | **89.38±1.88** |

## Algorithm comparison

*1) Comparison with classical feature-learning algorithms*

TABLE VI

Comparison of different feature-learning algorithms

(on the additional data sets)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | OF (%) | PCA (%) | LDA (%) | LPP (%) | Relief (%) | Lasso (%) | P\_value (%) | NE\_ESAE (%) |
| Pendigits | **98.13±0.05** | 98.07±0.13 | 97.87±0.23 | 97.97±0.11 | 97.82±0.23 | 98.13±0.05 | 92.87±1.46 | 84.21±0.83 |
| Vehicle | 80.35±1.31 | 82.34±1.05 | 82.55±0.92 | 81.77±0.89 | 78.16±0.19 | 80.781.05 | 75.39±0.89 | **89.53±1.96** |
| heart | 78.89±3.42 | 85.33±2.53 | 84.67±2.14 | 84.22±3.37 | 81.78±3.00 | 78.29±3.42 | 62.00±4.67 | **99.44±1.11** |
| Maxlettle | 84.62±3.92 | 88.62±2.79 | 88.62±4.16 | 86.77±4.43 | 85.85±5.03 | 85.23±4.16 | 76.62±2.75 | **96.92±1.09** |
| Wisconsin | 96.30±1.72 | 97.18±1.19 | 96.83±1.26 | 97.00±1.18 | 96.74±1.83 | 96.30±1.72 | 93.66±1.67 | **99.89±0.21** |
| PID | 70.40±2.74 | 72.34±1.98 | 75.78±3.49 | 69.67±5.54 | 73.29±4.51 | 70.39±2.74 | 73.91±3.55 | **89.38±1.88** |

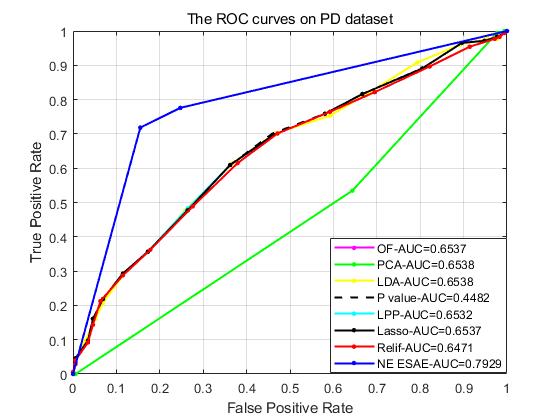


Fig. 1. Description of the ROC curves on PD.

*2) Comparison with* *state-of-the-art kernel feature methods*

TABLE VII

Comparison of different feature-learning algorithms

(on the additional data set)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | KPCA (%) | KLDA (%) | KDL (%) | KTL (%) | KLCTL (%) | NE\_ESAE (%) |
| Pendigits | **99.19±1.21** | 98.38±0.10 | 94.69±0.93 | 96.51±0.92 | 97.11±0.91 | 84.21±0.83 |
| Vehicle | 81.42±1.50 | 87.00±0.59 | 71.49±2.55 | 73.99±1.56 | 85.06±1.80 | **89.53±1.96** |
| heart | 85.11±3.30 | 81.11±2.36 | 80.79±0.47 | 84.07±0.93 | 86.94±0.87 | **99.44±1.11** |
| Maxlettle | 86.46±2.96 | 83.69±4.43 | 84.07±3.59 | 83.59±4.70 | 87.35±1.09 | **96.92±1.09** |
| Wisconsin | 97.47±1.31 | 94.01±2.03 | 94.36±0.13 | 95.23±1.16 | 96.94±0.67 | **99.89±0.21** |
| PID | 76.95±1.44 | 74.30±0.75 | 65.00±0.21 | 77.14±0.23 | 82.41±0.92 | **89.38±1.88** |

*3) Comparison with the representative stacked autoencoders*

TABLE VIII

Classification accuracy (mean ±variance) of different deep autoencoder classifiers

(on the additional data set)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | SAE (%) | SSAE (%) | SDSAE (%) | SPSAE (%) | ESGSAE\_FF (%) | GSTAE (%) | NE\_ESAE (%) |
| Vehicle | 67.30±3.33 | 70.00±2.99 | 72.00±2.25 | 74.76±2.93 | 81.91±0.42 | 79.71±2.93 | **89.53±1.96** |
| Pendigits | 89.64± 1.44 | 93.80± 0.51 | 94.58 ± 0.53 | 91.60 ± 0.57 | **98.00 ±0.12** | 93.53±0.77 | 84.21±0.83 |

## Parametric analysis

*1) Effects of classifier type*

TABLE IX

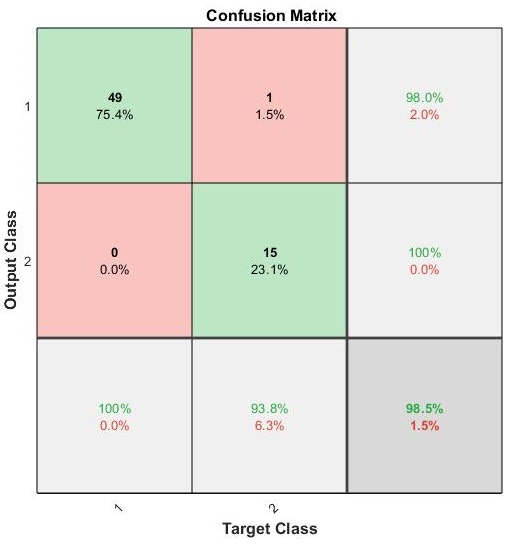
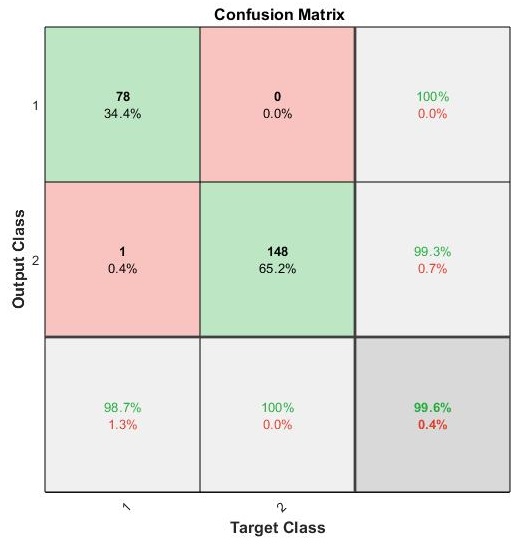
Classification accuracy (mean ±variance) of the proposed

algorithm with different classifiers

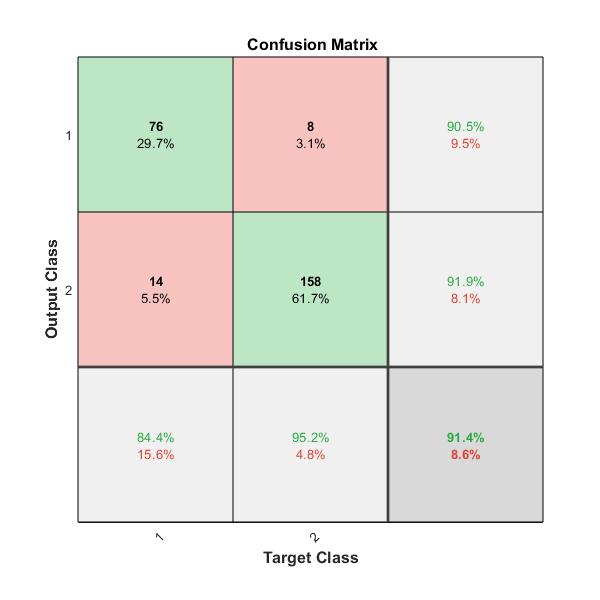
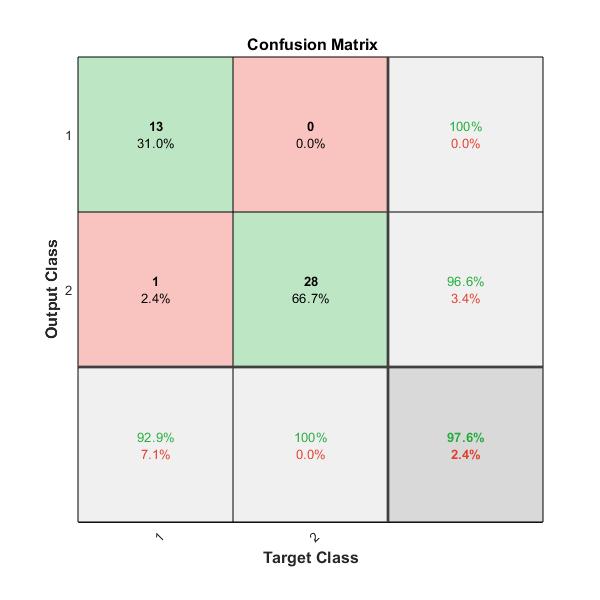
(on the additional data set)

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | SVM (%) | RF (%) | ELM (%) |
| Pendigits | 84.21±0.83 | 82.78±1.79 | **85.17±1.35** |
| Vehicle | **89.53±1.96** | 80.00±8.11 | 80.99±7.34 |
| heart | **99.44±1.11** | 98.89±1.28 | 98.33±2.13 |
| Maxlittle | 96.92±1.09 | **98.75±1.71** | **98.75±1.71** |
| Wisconsin | **99.89±0.21** | 99.41±0.51 | 99.71±0.51 |
| PID | 89.38±1.88 | **89.84±2.47** | 89.38±2.25 |

## Confusion matrix

****

(a) Maxlittle (b) Wisconsin

****

(c) PID (d) LSVT

Fig. 2. The confusion matrix under additional datasets

## Computational Complexity and Time Analysis

*1)* *Computational complexity:* The computational complexity of the proposed approach can be computed as a sum of four tasks: i) complexity of SPC, ii) complexity of ICMC and iii) complexity of ESAE.



where the , and  is the time complexity of SPC, ICMC and ESAE.

i) The computational complexity of SPC is , where  is the number of samples.

ii) The computational complexity of ICMC consists of four basic steps: IMC, update , update , and update . IMC is calculated as , where  is the number of iterations,  is the number of clustering clusters, and  is the number of samples. The complexity of computing and  is. The calculation of  involves feature decomposition and matrix multiplication, and the complexity is . Assuming that the number of iterations is , the total computational complexity of ICMC can be expressed as . It is worth noting that the computational complexity of the Gram matrix is not considered here, as it can be computed in advance without having to compute in ICMC.

iii) The computational complexity of ESAE is related to the number of hidden layer neurons, which are  respectively, so the computational complexity of ESAE is . At the same time, there is the back propagation part, whose computational complexity is , where  is the number of iterations,  is the number of samples, and  is the computational complexity of calculating the gradient of a single sample. So the total computational complexity of ESAE can be expressed as

Therefore, the complexity of the proposed approach is the sum of all given as



Maximize the computational complexity, the computational complexity approximation of NE\_ESAE can be expressed as



Table X is a comparison of computational complexity between our algorithm and other autoencoder models. It can be seen from the table that although the complexity of our algorithm is the highest, it is similar to that of other methods. The algorithm complexity of our method is mainly added to the ICMC part, but it is worth it, because ICMC has deeply mined the hierarchical information between samples and improved the accuracy of sample classification.

*2) Time Analysis:* Table XI shows the performance of running time and classification accuracy of additional data set in different methods. We can find that although our method does not have the shortest running time, it has the best sorting effect and the running time is at a medium level.

TABLE X

Comparison of complexity of different methods

|  |  |
| --- | --- |
| Method | Computational complexity |
| SAE |  |
| SSAE |  |
| SDSAE |  |
| SPSAE |  |
| ESGSAE\_FF |  |
| GSTAE |  |
| NE\_ESAE |  |

TABLE XI

Comparison of training time and classification accuracy of different methods in seconds.

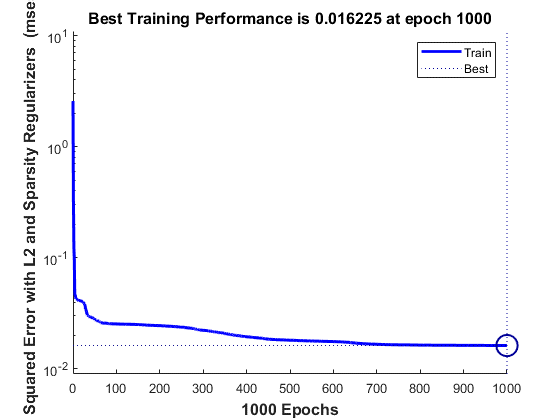
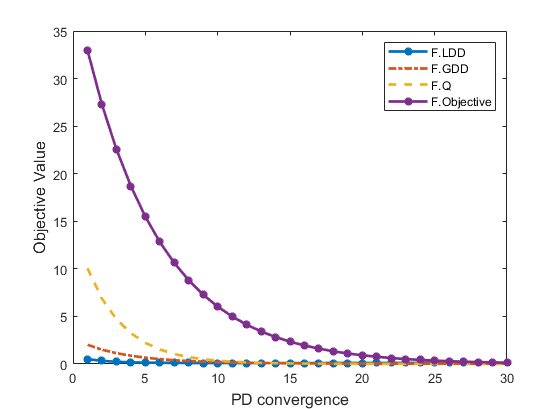
(on the additional data set)

|  |  |
| --- | --- |
| Method | Vehicle (%) |
| SAE | 34.88s (67.30%) |
| SSAE | 50.09s (70.00%) |
| SDSAE | 43.21s (72.00%) |
| SPSAE | 150.9s (74.76%) |
| ESGSAE\_FF | 130.8s (81.91%) |
| GSTAE | 30.22s (79.71%) |
| NE\_ESAE | 125.7s (89.53%) |

## Model Visualization and Convergence Analysis

The convergence curves of NE\_ESAE on additional data sets are shown in Figure 3.

In ICM experiment, the number of iterations is set to 30, and the changes of the objective function were shown in Figure 3 (a). In ESAE experiment, the number of iterations was set to 1000, and the changes of the objective function were shown in Figure 3 (b). By running the algorithm, the objective function is reduced to a constant value on the dataset PD after several iterations.



(a) Convergence of ICM on PD (b) Convergence of ICM on PD

Fig. 3. Convergence of our algorithm on PD