IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 27, 2020, accepted July 14, 2020, date of publication July 17, 2020, date of current version July 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3010096

Gas Path Fault Diagnosis of Aeroengine Based on Soft Square Pinball Loss ELM

YUYUAN CAO, BOWEN ZHANG, HUAWEI WANG, AND YU BAI
College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210000, China

Corresponding author: Yuyuan Cao (caoyuyuan@nuaa.edu.cn)

ABSTRACT The ELM constructed based on the least squares loss function and ±1 label has poor generalization in the classification of data containing noise. The introduction of square pinball loss function can improve the robustness of ELM. However, the algorithm based on the squared loss function and the ±1 label imposes a margin of 1 for all training samples. At the same time, due to the unbounded nature of the loss function, the generalization of the algorithm in the classification problem is reduced. This paper proposes a soft threshold square pinball loss (SSP-Loss) function. This function can set more flexible thresholds for training samples while maintaining the robustness of the square pinball loss function. The soft-threshold square pinball loss function can approximate the bounded loss function in stages to further improve the classification performance of the algorithm. The performance of ELM based on the soft pinball loss function on several benchmark data sets proves the effectiveness of our proposed algorithm. More importantly, the excellent robustness and classification performance of the algorithm is very suitable for aeroengine gas path fault diagnosis, and is expected to become its candidate technology.

INDEX TERMS Extreme learning machine, fault diagnosis, aircraft engine, machine learning.

I. INTRODUCTION

Due to the harsh working environment, large number of parts and complicated internal structure of the aeroengine, performance degradation inevitably occurs during operation, so it's important to take some measures to get these signs of degradation in advance to get the rest of the engine's life, hence the field of engine performance parameter prediction appeared. The physical model of the engine is very complicated, and the correlation between several parts is complicated, so it is very difficult to establish the model to predict the parameters. In recent years, with the development of machine learning methods, data-driven prediction methods have attracted more and more researchers' attention. Data-based methods do not require complex research models, and the accuracy of their predictions depends heavily on historical data. In the field of engine health detection, more and more data-based methods have been applied [1]–[5], these studies also prove that the correct use of data-based methods can effectively improve the accuracy of diagnosis.

Extreme learning machine (ELM) is a neural network training method proposed by Huang [6] in 2006. Since it was proposed, it has been widely used in classification, prediction and other fields due to its fast learning speed and good performance. However, two important problems restrict the application of ELM in more complex practical scenarios. One is its sensitivity to noise data, that is, poor robustness; the other is its poor generalization. These deficiencies are particularly fatal in the field of aircraft engine fault diagnosis. The data collected by sensors of aeroengines often contain a lot of noise with outliers, some of these abnormal data can be eliminated through preprocessing, but others cannot be eliminated. Therefore, the algorithm that can be used in the field of aircraft engine fault diagnosis must have excellent robustness. At the same time, in view of the importance of fault diagnosis for aviation safety, the algorithm in this field has higher requirements for classification accuracy. It is of great research significance to have good classification performance under the premise of stickiness.

As an algorithm for learning based on square loss, ELM is extremely sensitive to noise data, and its poor robustness restricts its application in many scenarios with noise data. In order to solve the shortage of the square loss function, many scholars have made a breakthrough in the study of the loss function, and proposed many new loss functions with good robustness. For example, the Hinge loss-based

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

SVM (Support Vector Machine) [7], [8] which adopts the truncation strategy to enhance the robustness of the function; SVR(Support Vector Regression) based on insensitive ε loss [9], [10]; the least-squares based truncation square loss SVM [11]; multiclass capped ℓp-norm SVM [12] *et al.* The above algorithms have achieved performance that can obtain better training effects under the interference of noise by modifying the loss function to some extent. The two typical ones are correntropy-induced loss (C-loss) function and pinball loss (P-loss) function. C-loss function was first proposed by Singh [13] to be applied in pattern classification, the neural network based on C-loss function shows better robustness and classification performance; Xu *et al.* [14] then applies the C-loss function to the study of the kernel method and proposes a robust C-loss kernel classifier; Chen *et al.* [15] made a further expansion, expanding the C-loss function into the case of central variability, in this article, he first mentioned the introduction of C-loss function into ELM for the analysis of regression problem; Zhao *et al.* [16] set the specification parameter of C-loss as 1 and introduced it into ELM to systematically deduce the solution process of ELM based on C-loss on the regression problem. The core idea of P-loss is to learn based on quantile distance [17]–[20], Huang *et al.* [18] first introduced this idea into SVM. Compared with the C-loss function, the research on the P-loss function is more mature. In recent years, many scholars have kept studying it [21]–[27] and made some progress. Wang and Ding [28] proposed a square P-loss (SP-loss) function based on the existing progress in 2019, and introduced the loss function into the learning of ELM, and obtained SPELM, showing superior robustness on the regression problem. However, whether it is C-loss function or SP-loss function, there is still a lack of research on type ELM. Based on these two new loss functions, we study their performance in classification problems.

Another problem that restricts ELM in complex practical applications is its generalization. It is not difficult to find through research that ELM is a square-based learning algorithm, which on regression problems will not restrict its generalization. However, when learning classification problems, no matter which loss function is used, ± 1 labels are usually used as the output to learn the two classification problems. The training method based on square learning will force the margins of all training samples to approach 1, which on the one hand does not meet the maximum margin learning idea similar to SVM in statistics, and the unbounded square loss function greatly reduces its generalization [29], on the other hand, will increase the risk of algorithm overfitting. The ELM algorithm based on square pinball loss function and C-loss function is proposed for two kinds of classification.

In order to improve the limitations of ELM robustness and versatility, we propose an ELM based on the soft pinball loss function (SSPELM). By introducing the square pinball loss function to improve the robustness of ELM, and by introducing a relatively flexible margin to make the square loss function approximate the bounded function in stages to improve its generalization ability, the main contributions of this paper can be summarized as follows:

1) The ELM algorithm based on the two classifications of square pinball loss function and C loss function is studied;

2) An ELM algorithm based on soft threshold pinball loss function is proposed to solve the classification problem;

3) The algorithm is experimentally verified on several benchmark data sets, and the experimental results prove the superiority of the algorithm.

4) The new loss function ELM is used for fault diagnosis of aero-engine air circuits, and has achieved good diagnostic results.

II. RELATED WORK

A. EXTREME LEARNING MACHINE (ELM)

In this section, we briefly review ELM. Assume the data to be trained is $(\mathbf{x}, t)$, dimension for $N \times (m+1)$, $\boldsymbol{x}$ is the sample input data, t is the sample output data, N is the number of samples, m is the dimension of the sample input data, that is, the number of input features, the number of neurons in the hidden layer is M, the connection weight of the input layer of ELM to the hidden layer is $\mathbf{A}$, the dimension is $M \times m$, the threshold value is $\mathbf{B}$, and the dimension is $M \times 1$, the activation function of hidden layer is expressed as h($\mathbf{x}$), and the output of hidden layer can be expressed as follows:

$$\boldsymbol{H} = h(\boldsymbol{A}\boldsymbol{x}^T + \boldsymbol{B}) \tag{1}$$

Next, to solve the connection weight of the hidden layer to the output layer β. In the network training stage, the network output is known, that is, $\boldsymbol{T}$ is known, and β is to be solved. Solving β is equivalent to solving the optimal solution of the following loss function:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{T} - \boldsymbol{t}\|_2^2 = \min_{\boldsymbol{\beta}} \|\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{t}\|_2^2 \tag{2}$$

According to the least square method, it is not difficult to obtain the final solution of β as follows:

$$\boldsymbol{\beta} = \boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{3}$$

where $\boldsymbol{H}^{+}$ Represents the Moore–Penrose generalized inverse of $\boldsymbol{H}$.

B. REGULARIZED EXTREME LEARNING MACHINE (RELM)

In this section, we review RELM [30]. According to formula (2), the solution of ELM is a process to minimize the empirical risk. But in some cases, this solution presents the problem of overfitting, scholar Deng introduced structural risk to avoid overfitting, and updated the loss function of ELM as follows:

$$\min_{\boldsymbol{\beta}} \frac{1}{2}\|\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{t}\|_2^2 + \frac{C}{2}||\boldsymbol{\beta}||_2^2 \tag{4}$$

C is the regularization parameter. According to the least square method, the final solution of equation (4) is:

$$\boldsymbol{\beta} = \begin{cases} (\boldsymbol{H}^T\boldsymbol{H} + \dfrac{I}{C})^{-1}\boldsymbol{H}^T\boldsymbol{t}, & N \geq M \\ \boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{H}^T + \dfrac{I}{C})^{-1}\boldsymbol{t}, & N < M \end{cases} \tag{5}$$

$\boldsymbol{I}$ is the unit matrix whose dimension is the number of neurons in the hidden layer. In general, the solution obtained in equation (5) has better performance than that obtained in equation (4), with more-sparse and does not increase too much in computational complexity, which is a good choice in practical applications.

III. ELM BASED ON THE NEW LOSS FUNCTION

ELM and RELM both adopt the weight solving algorithm based on square loss, although the algorithm based on the square loss function is a very common method, it is limited in many complex practical applications due to its sensitivity to noise. In order to improve the robustness of ELM, we introduce the P-loss function and C-loss function to improve ELM in this section. The P-loss function is a loss function proposed in recent years and has made some progress in the application of SVM. C-loss function is a loss function developed based on entropy theory. ELM based on C-loss shows better regression effect and robustness than ELM and RELM in the regression problem. Studies on the classification of C-loss ELM are still lacking. The square pinball loss function (l-sp) and the C-loss loss function (l-c) are as follows:

$$l_{sp}(e_i) = \begin{cases} pe_i^2, & e_i \geq 0 \\ (1-p)e_i^2, & e_i < 0 \end{cases} \tag{6}$$

$$l_C(e_i) = 1 - \exp(-\frac{e_i^2}{2\sigma^2}) \tag{7}$$

In equation (6), $0 \leq p \leq 1$, based on equation (6), we can obtain the loss function of ELM based on P-loss, and the solution objective becomes minimized under certain conditions:

$$\begin{aligned} &\min_{\boldsymbol{\beta}, e_i} \frac{1}{2}\|\boldsymbol{\beta}\|_2^2 + \frac{C}{2}\sum_{i=1}^{N} l_{sp}(e_i) \\ &S.t.\ h(\boldsymbol{x}_i)\boldsymbol{\beta} = t_i - e_i, \quad i = 1, 2, \ldots N \end{aligned} \tag{8}$$

The solution of equation (8) is not complicated. Lagrange multiplier method can be used to better solve this problem. We can get:

$$\boldsymbol{\beta} = \begin{cases} (\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H} + \frac{I}{C})^{-1}\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{t}, & N \geq M \\ \boldsymbol{H}^T(\boldsymbol{W}\boldsymbol{H}^T\boldsymbol{H} + \frac{I}{C})^{-1}\boldsymbol{W}\boldsymbol{t}, & N < M \end{cases} \tag{9}$$

where $\boldsymbol{W}$ is a diagonal matrix, which can be evaluated as follows:

$$\boldsymbol{W} = diag(w_1^{sp}, w_2^{sp}, \ldots w_N^{sp}) \tag{10}$$

$$w_i^{sp} = \frac{\frac{1}{2}\frac{\partial l_{sp}(e_i)}{\partial e_i}}{e_i} = \begin{cases} p, & e_i \geq 0, \\ 1-p, & e_i < 0 \end{cases} \tag{11}$$

By solving the above equation iteratively, the final convergent solution can be obtained. The iteration stop condition can be set as the difference between $\boldsymbol{\beta}$ obtained by two solutions is less than the set threshold ε, or the iteration times reach the set maximum itMax. We can summarize the algorithm flow of SPELM as follows:

Algorithm 1 SPELM

Input: Training sample set $\{(\boldsymbol{x}_i, t_i)\}_{i=1}^N$; activation function;
parameters: $p, C, M, \varepsilon, itMax$.
Output: $\boldsymbol{\beta}$
Initialize:
calculate the hidden nodes output matrix $\boldsymbol{H}$;
let $k = 0$, calculate β according to (5);
Denote $w_i^{sp} = \begin{cases} p, & e_i \geq 0, \\ 1-p, & e_i < 0, \end{cases}$ $\boldsymbol{W} = diag(w_1^{sp}, w_2^{sp}, \ldots w_N^{sp})$;
While $k < itMax$ and $\left\|\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1}\right\| > \varepsilon$
calculate $\boldsymbol{\beta}$ according to (9);
calculate $\boldsymbol{W}$ according to (10), (11);
$k = k + 1$;
End while

The loss function of CELM is based on equation (8), replace the square pinball loss function (l-sp) in the equation with the C-loss function (l-c). Since non-convex loss function is used, semi-optimal algorithm is needed to solve the problem. There is a detailed derivation process in literature [31], which will not be described in this paper. However, for different parts of classification problem and regression problem, error e is determined in a different way. The e calculation of classification problem is carried out according to the following formula:

$$e_i = 1 - t_i h(\boldsymbol{x}_i)\boldsymbol{\beta}, \quad i = 1, 2, \ldots N \tag{12}$$

CELM's algorithm is summarized as follows:

Algorithm 2 CELM

Input: Training sample set $\{(\boldsymbol{x}_i, t_i)\}_{i=1}^N$; activation function;
parameters: $\sigma, C, M, \varepsilon, itMax$.
Output: $\boldsymbol{\beta}$
Initialize:
calculate the hidden nodes output matrix $\boldsymbol{H}$;
let $k = 0$, calculate $\boldsymbol{\beta}$ according to (5);
While $k < itMax$ and $\left\|\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1}\right\| > \varepsilon$
$e_i = 1 - t_i h(x_i)\boldsymbol{\beta}, v_i^k = -\exp\{-\frac{e_i^2}{2\sigma^2}\}, i = 1, 2, \ldots N$;
$\boldsymbol{\Omega} = diag\{-v_1^k, -v_2^k, \ldots - v_N^k\}$;
$\boldsymbol{\beta}^k = \begin{cases} (\boldsymbol{H}^T\boldsymbol{\Omega}\boldsymbol{H} + \frac{I}{C})^{-1}\boldsymbol{H}^T\boldsymbol{\Omega}\boldsymbol{t}, & N \geq M \\ \boldsymbol{H}^T(\boldsymbol{\Omega}\boldsymbol{H}^T\boldsymbol{H} + \frac{I}{C})^{-1}\boldsymbol{\Omega}\boldsymbol{t}, & N < M \end{cases}$
$k = k + 1$;
End while

IV. ELM BASED ON SOFT SQUARE PINBALL LOSS

Both ELM and RELM based on least square loss, or SPELM based on pinball loss and CELM based on C-loss function, that make $\boldsymbol{H\beta}$ approach t, which is reasonable in the regression problem, because t is a continuous value at this time. When we do binary classification problems, usually use the label ± 1 to represent t, at this point, the discriminant function is different from the regression problem, and the following equation is generally used for the discrimination of ± 1 category:

$$f(\boldsymbol{x}_i) = sign(h(\boldsymbol{x}_i)\boldsymbol{\beta}), \quad i = 1, 2, \ldots N \tag{13}$$

When the value in parentheses of equation (13) is not less than 0, $f(x) = +1$, otherwise it's -1, As for the loss function of ELM mentioned above, we can find that: in the classification problem with ± 1 label, if the algorithm is based on the square loss function, the square of the network output of all samples will approach 1. However, according to formula (13), it can be seen that for $+1$ label samples, as long as the network output is greater than 0,and for -1 label samples, the network output is less than 0, and it is not necessary to force that the square of all sample network output is equal to 1. This kind of hard threshold reduces the generalization of the algorithm and does not conform to the maximum margin principle of SVM algorithm. In this paper, we introduce a more flexible threshold determination method, based on this threshold determination method, the soft square pinball loss function ELM (SSPELM) is proposed. The loss function of SSPELM can be expressed as follows:

$$\begin{aligned} &\min \frac{1}{2}\|\boldsymbol{\beta}\|_2^2 + \frac{C}{2}\sum_{i=1}^{N} l_{sp}(e_i), \quad i = 1, 2, \ldots N \\ &S.t. \boldsymbol{H\beta} + e = \boldsymbol{t} \odot \delta \end{aligned} \tag{14}$$

In formula, $\odot$ represents the Hadamard product, $\delta = [\delta_1, \delta_2, \ldots \delta_{\mathrm{N}}]^{\mathrm{T}}$, δ represents the threshold values of some columns introduced, for each training sample there is a δ corresponding to it, SSPELM is SPELM when all $\delta = 1$, so SPELM is a special case of SSPELM.

The idea of solving equation (14) is to adopt the two-step iterative solution method. In the first step, we first assume that δ is known after the Kth iteration, and then solve for $\boldsymbol{\beta}$. At this time, the solution idea is the same as SPELM, by using Lagrange multiplier method, we can get:

$$\boldsymbol{\beta}^{k+1} = \begin{cases} (\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H} + \dfrac{I}{C})^{-1}\boldsymbol{H}^T\boldsymbol{W}(\boldsymbol{t} \odot \delta^k), & N \geq M \\ \boldsymbol{H}^T(\boldsymbol{W}\boldsymbol{H}^T\boldsymbol{H} + \dfrac{I}{C})^{-1}\boldsymbol{W}(\boldsymbol{t} \odot \delta^k), & N < M \end{cases} \tag{15}$$

The determination of W in the equation is still calculated according to equations (10) and (11). Slightly different from SPELM, e is calculated when calculating W. For the i-th training sample, the corresponding e_i is calculated as follows:

$$e_i = t_i\delta_i - h(\boldsymbol{x}_i)\boldsymbol{\beta}, \quad i = 1, 2, \ldots N \tag{16}$$

In the second step, given the β of k + 1 iteration, solve δ:

At this time, the network prediction value of each training sample can be calculated with determination. We define a variable:

$$s_i^{k+1} = t_i f^{k+1}(\boldsymbol{x}_i) = t_i h(\boldsymbol{x}_i)\boldsymbol{\beta}^{k+1}, \quad i = 1, 2, \ldots N \tag{17}$$

Then, we take the value calculated in equation (17) as the classification standard, and update δ according to certain rules [32]:

(1) If $s_{\mathrm{i}}^{\mathrm{k+1}} > 1$,this situation indicates that the i-th training sample has been correctly classified and the output of the network is greater than 1, which far meets the margin requirement of equation (13), therefore, the threshold corresponding to the i-th sample of the next iteration can be set as:

$$\delta_i^{k+1} = s_i^{k+1} \tag{18}$$

(2) If $0 \leqslant s_{\mathrm{i}}^{\mathrm{k+1}} \leqslant 1$, in this situation, it means that although the i-th training sample has been correctly classified, the network output is between 0 and 1, and the margin is slightly insufficient. Therefore, the threshold corresponding to the i-th sample in the next iteration can be set as:

$$\delta_i^{k+1} = 1 \tag{19}$$

(3) If $s_{\mathrm{i}}^{\mathrm{k+1}} < 0$, in this situation, it means that the i-th training sample is misclassified, and the threshold corresponding to the i-th sample in the next iteration can be set as:

$$\delta_i^{k+1} = s_i^{k+1} + 1 \tag{20}$$

According to equations (18) to (20) above, we can obtain the threshold vector δ corresponding to the next iteration. The above three expressions can be summarized as the following update rules:

$$\delta_i^{k+1} = s_i^{k+1} + \max\{0, \min\{1, 1 - s_i^{k+1}\}\} \tag{21}$$

In the first iteration, we initialized δ as all 1 vectors, which $\boldsymbol{\beta}$ can be initialized according to formula (5). The algorithm of SSPELM can be summarized as follows:

Algorithm 3 SSPELM

Input: Training sample set $\{(x_i, t_i)\}_{i=1}^N$; activation function; parameters: $p, C, M, \varepsilon, itMax$.
Output: $\boldsymbol{\beta}$
Initialize:
calculate the hidden nodes output matrix $\boldsymbol{H}$;
let $k = 0$, $\delta = 1$, calculate β according to (5);
$e_i = t_i\delta_i - h(x_i)\boldsymbol{\beta}$;
Denote $w_i^{sp} = \begin{cases} p, & e_i \geq 0, \\ 1-p, & e_i < 0, \end{cases}$ $\boldsymbol{W} = diag(w_1^{sp}, w_2^{sp}, \ldots w_N^{sp})$;
While $k < itMax$ and $\left\|\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1}\right\| > \varepsilon$
calculate $\boldsymbol{\beta}$ according to (15);
$e_i = t_i\delta_i - h(x_i)\boldsymbol{\beta}$;
calculate $\boldsymbol{W}$ according to (10), (11);
calculate δ according to (21);
$k = k + 1$;
End while

TABLE 1. Data set settings.

Data sets	Features	Class	Samples
German	20	2	1000
Breast Cancer	10	2	699
Biodeg	41	2	1055
Audit_Risk	26	2	776
Dermatology	34	6	366
Data_Banknote_ Authentication	4	2	1372

V. NUMERICAL EXPERIMENTS

In order to verify the effectiveness of the algorithm, we selected a typical UCI [33] benchmark data set for numerical experiments, including ELM, RELM, CELM, SPELM and SSPELM. The selected data set information is as TABLE 1:

For each data set, we randomly divided it according to the proportion of 60% training samples and 40% test samples, and conducted 10 experiments for each data set. For the model parameters in RELM, CELM, SPELM and SSPELM, we confirmed the training samples in the first randomized experiment with K-CV [34] verification. Take K as 5, the values of parameters σ and C are from $\{2^{-5}, 2^{-4}\ldots 2^{4}, 2^{5}\}$, the value of p is determined from $\{0, 0.1, 0.2\ldots 0.9, 1\}$, the number of neurons in the hidden layer was set as 20, itMax was set as 100, and ε was set as 10^{-10}. The activation function takes sigmoid function and Sine function respectively for the experiment. During the experiment, the input of all data is normalized to [0, 1], and the output is normalized to $\{-1, +1\}$. For the case of multi-classification, this paper adopts the "one-to-many" approach, that is, one category of multiple categories is taken as one class at a time, and all other categories are taken as another class for training. After the training of multiple classifiers, "voting" is adopted to determine the category of test samples.

All experiments were conducted in the following test environment: Intel(R)Core(TM)i5-8265U CPU @1.60 GHz,

TABLE 2. The experimental results of sigmoid activation function are noiseless.

Dataset	Method	(C, σ, p)	ACC (%)	MCC (%)	F_1 (%)
German	ELM	(/, /, /)	74.05±2.21	**31.11±5.08**	83.24±0.60
	RELM	(2^1, /, /)	**74.25±1.49**	30.40±5.58	83.20±0.65
	CELM	(2^5, 2^1, /)	74.10±1.88	30.67±3.34	83.29±1.00
	SPELM	(2^2, /, 0.5)	**74.25±1.49**	30.57±5.47	83.29±0.80
	SSPELM	(2^1, /, 0.5)	73.85±1.38	29.23±7.23	**83.53±1.01**
Breast Cancer	ELM	(/, /, /)	96.20±1.06	91.02±2.93	96.87±1.19
	RELM	(2^2, /, /)	96.13±0.93	90.78±5.05	96.77±1.93
	CELM	(2^4, 2^1, /)	95.98±0.96	90.80±5.53	96.77±2.10
	SPELM	(2^0, /, 0.1)	97.06±1.20	92.79±1.89	97.37±0.80
	SSPELM	(2^4, /, 0.2)	**97.13±0.36**	**94.01±2.75**	**97.81±1.10**
Biodeg	ELM	(/, /, /)	76.59±4.16	**51.59±7.04**	63.82±11.08
	RELM	(2^0, /, /)	76.49±4.29	49.62±8.80	62.17±13.03
	CELM	(2^4, 2^{-2}, /)	76.45±4.32	49.38±9.35	61.85±13.69
	SPELM	(2^{-2}, /, 0.6)	78.15±3.35	46.33±8.62	62.20±9.80
	SSPELM	(2^{-1}, /, 0.6)	**78.96±3.76**	49.05±6.24	**65.14±7.01**
Audit_Risk	ELM	(/, /, /)	93.87±2.40	91.62±2.06	94.80±1.31
	RELM	(2^0, /, /)	94.32±1.04	91.50±0.78	94.69±0.66
	CELM	(2^5, 2^2, /)	94.45±1.50	91.50±0.72	94.71±0.54
	SPELM	(2^1, /, 0.6)	95.16±1.09	92.15±1.13	95.21±0.81
	SSPELM	(2^5, /, 0.5)	**96.00±1.20**	**93.10±0.45**	**95.78±0.18**
Dermatology	ELM	(/, /, /)	97.92±1.05	90.22±3.87	91.34±1.05
	RELM	(2^0, /, /)	97.12±1.20	88.77±4.64	89.36±1.20
	CELM	(2^4, 2^2, /)	96.99±1.23	88.70±6.18	89.24±1.23
	SPELM	(2^1, /, 0.5)	97.12±1.20	88.77±4.64	89.36±1.20
	SSPELM	(2^4, /, 0.5)	**98.24±0.79**	**92.20±2.81**	**93.22±0.79**
Data_Banknote _Authenticatio n	ELM	(/, /, /)	**99.38±0.60**	**97.53±1.28**	**98.87±0.60**
	RELM	(2^4, /, /)	98.00±0.72	94.71±1.44	97.50±0.72
	CELM	(2^4, 2^0, /)	97.81±0.85	94.37±1.80	97.33±0.90
	SPELM	(2^5, /, 0.6)	98.43±0.74	95.41±1.34	97.85±0.65
	SSPELM	(2^4, /, 0.6)	98.76±0.33	95.65±1.04	97.99±0.51

TABLE 3. The experimental results of sine activation function are noiseless.

Dataset	Method	（C， σ， p）	ACC （%）	MCC （%）	F_1 （%）
German	ELM	（/， /， /）	72.90±2.75	29.50±3.74	81.81±0.51
	RELM	（2^5， /， /）	72.90±2.75	29.50±3.74	81.81±0.51
	CELM	（2^2， 2^1， /）	73.00±2.44	28.21±2.65	81.77±1.06
	SPELM	（2^2， /， 0.5）	72.90±2.75	29.31±3.31	82.07±0.50
	SSPELM	（2^2， /， 0.5）	**73.15±2.68**	**31.01±4.52**	**82.20±0.64**
Breast Cancer	ELM	（/， /， /）	95.55±1.06	92.08±2.37	97.27±0.79
	RELM	（2^{-1}， /， /）	95.70±1.05	92.24±2.03	97.32±0.66
	CELM	（2^3， 2^3， /）	95.63±1.53	91.57±2.20	97.12±0.72
	SPELM	（2^5， /， 0.5）	95.63±1.12	92.26±1.56	97.32±0.52
	SSPELM	（2^1， /， 0.4）	**96.18±0.86**	**92.94±0.65**	**97.53±0.16**
Biodeg	ELM	（/， /， /）	75.88±3.39	**54.50±6.19**	70.62±3.56
	RELM	（2^{-3}， /， /）	76.16±3.58	53.36±6.75	69.68±3.27
	CELM	（2^1， 2^2， /）	76.16±3.44	53.53±6.55	69.79±3.34
	SPELM	（2^{-4}， /， 0.6）	**76.87±5.84**	52.33±9.11	69.95±5.58
	SSPELM	（2^{-2}， /， 0.6）	75.58±5.32	53.36±12.12	**70.66±7.35**
Audit_Risk	ELM	（/， /， /）	92.52±1.92	90.41±0.55	93.93±0.37
	RELM	（2^{-1}， /， /）	93.29±1.82	**91.47±1.00**	94.66±0.70
	CELM	（2^5， 2^3， /）	93.36±1.73	**91.47±1.00**	94.66±0.70
	SPELM	（2^{-4}， /， 0.7）	92.13±1.67	86.42±4.60	91.84±2.75
	SSPELM	（2^{-2}， /， 0.5）	**94.18±2.25**	91.43±2.07	**94.69±1.42**
Dermatology	ELM	（/， /， /）	95.76±1.93	83.09±7.42	84.81±7.15
	RELM	（2^1， /， /）	**95.80±1.89**	82.76±7.98	84.32±7.84
	CELM	（2^5， 2^2， /）	95.78±1.83	82.70±8.10	84.28±7.93
	SPELM	（2^{-2}， /， 0.5）	**95.80±1.89**	82.76±7.98	84.32±7.84
	SSPELM	（2^0， /， 0.5）	95.69±2.06	**84.40±6.44**	**86.71±5.78**
Data_Banknote_Authentication	ELM	（/， /， /）	**99.53±0.36**	99.49±0.80	99.77±0.37
	RELM	（2^5， /， /）	98.69±0.79	98.68±1.47	99.41±0.66
	CELM	（2^5， 2^2， /）	98.65±0.53	97.82±1.46	99.00±0.65
	SPELM	（2^5， /， 0.7）	99.31±0.58	**99.93±0.17**	**99.97±0.08**
	SSPELM	（2^5， /， 0.4）	99.38±0.71	99.28±1.25	99.66±0.59

8.00GB memory, and Windows 10 operating system in MATLAB R2016a environment. The two activation functions are calculated as follows:

$$\text{Sigmoid: } h(x) = \frac{1}{1+\exp[-(Ax^T+B)]} \tag{22}$$

$$\text{Sine: } h(x) = \sin(Ax^T+B) \tag{23}$$

The evaluation indexes of classification effect are as follows:

$$ACC = \frac{TP+TN}{TP+FN+TN+FP} \tag{24}$$

$$F_1 = \frac{2TP}{2TP+FP+FN} \tag{25}$$

$$MCC = \frac{TP\times TN - FP\times FN}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}} \tag{26}$$

ACC means accuracy, F1 score is the harmonic mean of precision and sensitivity, MCC means Matthews correlation coefficient. A sample labeled +1 is called a positive sample, a sample labeled −1 is called a negative sample, in equations (24) to (26), TP represents the number of positive samples correctly classified, TN represents the number of negative samples correctly classified, FP represents the number of positive samples incorrectly classified, and FN represents the number of negative samples incorrectly classified. The larger of those index, the better.

The numerical experiments performed on six data sets are briefly described below:

(1) The experiment of classification data set was carried out without adding any noise data: According to the proportion of 60% of the training samples and 40% of the test samples, the model parameters are selected from the values determined by K-CV for the training samples in the first randomized

TABLE 4. Add noise to experimental results under sigmoid activation function.

Dataset	Method	（C，σ，p）	ACC（%）	MCC（%）	F_1（%）
German	ELM	（/，/，/）	67.85±5.81	**29.64±8.28**	76.85±4.61
	RELM	（2^1，/，/）	71.65±4.27	28.45±8.90	79.96±1.03
	CELM	（2^5，2^1，/）	70.80±4.42	27.20±11.23	79.39±2.29
	SPELM	（2^1，/，0.4）	71.65±4.27	28.45±8.90	79.96±1.03
	SSPELM	（2^1，/，0.5）	**74.50±2.23**	26.15±14.61	**82.53±1.70**
Breast Cancer	ELM	（/，/，/）	89.11±3.11	79.92±2.61	93.12±1.42
	RELM	（2^2，/，/）	90.25±2.27	79.82±5.80	93.16±1.98
	CELM	（2^4，2^1，/）	90.25±2.39	79.10±5.47	92.91±1.98
	SPELM	（2^4，/，0.3）	96.49±1.42	79.61±5.07	93.40±1.68
	SSPELM	（2^4，/，0.2）	**96.70±0.59**	**85.08±4.36**	**94.88±1.53**
Biodeg	ELM	（/，/，/）	56.07±8.49	17.97±10.63	48.84±10.46
	RELM	（2^0，/，/）	55.21±5.06	21.70±16.24	48.97±17.99
	CELM	（2^4，2^{-2}，/）	55.88±5.27	21.67±16.96	48.34±19.75
	SPELM	（2^{-1}，/，0.5）	62.37±8.83	22.94±16.46	47.61±21.69
	SSPELM	（2^{-1}，/，0.6）	**75.59±3.89**	**23.14±15.01**	**48.93±18.93**
Audit_Risk	ELM	（/，/，/）	77.74±5.03	61.57±12.45	66.37±13.77
	RELM	（2^0，/，/）	82.71±5.91	59.78±12.55	64.33±14.88
	CELM	（2^5，2^2，/）	81.23±5.07	60.11±12.94	64.63±15.18
	SPELM	（2^1，/，0.6）	**84.32±6.07**	**63.93±12.42**	**69.16±14.09**
	SSPELM	（2^1，/，0.5）	79.80±14.03	60.14±12.52	64.75±14.64
Dermatology	ELM	（/，/，/）	79.41±9.90	73.82±7.93	75.91±9.90
	RELM	（2^0，/，/）	90.82±4.45	**78.36±7.09**	**79.54±5.45**
	CELM	（2^4，2^2，/）	**91.07±5.14**	76.31±7.85	77.31±5.14
	SPELM	（2^1，/，0.5）	88.75±7.34	**78.36±7.09**	**79.54±7.34**
	SSPELM	（2^1，/，0.6）	90.50±5.00	75.50±6.70	77.04±5.00
Data_Banknote_Authentication	ELM	（/，/，/）	73.18±9.31	65.15±8.43	86.20±2.92
	RELM	（2^4，/，/）	75.05±4.87	**67.77±4.56**	**87.09±1.72**
	CELM	（2^4，2^0，/）	**75.40±6.99**	66.55±4.66	86.67±1.57
	SPELM	（2^4，/，0.5）	69.34±5.63	54.87±7.45	82.56±2.52
	SSPELM	（2^4，/，0.6）	72.08±5.22	54.27±7.52	82.35±2.52

experiment. The results of the three indicators and the model parameters of the algorithm obtained under the two activation functions are shown in Table 2 and Table 3;

(2) Add noise data subject to N (0,0.5) distribution to the input characteristics of each dataset: In the input characteristics of the original data, gaussian white noise obeying a certain distribution was added, and then the classification experiment was carried out. Other Settings remained unchanged. The results of the three indicators obtained are shown in Table 4 and Table 5;

(3) Consider the computational complexity of the algorithm: The iterations of CELM, SPELM and SSPELM, as well as the training time and test time of the five algorithms, are recorded. The indicators related to the calculation time consumption are shown in Table 6.

By analyzing the experimental results obtained in Table 2 and Table 3, the following conclusions can be obtained: for the algorithm SSPELM proposed in this paper, four of the six data sets have the optimal classification performance, it is proved that SSPLEM algorithm is universal and superior to the other four algorithms; The six data sets contain binary and multi-classification situations. The index results in the table prove that SSPELM has excellent performance in both binary and multi-classification situations; In the case of binary classification, the input characteristic dimension of the data set increases successively from 4 to 40. However, in the range of this extensive input characteristic dimension, SSPELM performs better than the algorithm of ELM, RELM, CELM and SPELM, which proves the universality of the algorithm from another level; The sample size of the data set was from 300 to 1300, and the algorithm obtained good predictive performance by learning under different training samples. The number of hidden layer neurons has a great impact on the performance of ELM. We sequentially increased the number of hidden layer neurons of the five algorithms from 10 to 100, and verified the classification performance of several algorithms on several data sets. Figures 1 to 6 show the performance of several algorithms in one experiment.

TABLE 5. Add noise to experimental results under sine activation function.

Dataset	Method	（C，σ，p）	ACC（%）	MCC（%）	F1（%）
German	ELM	（/，/，/）	64.70±7.06	16.17±10.32	71.67±5.54
	RELM	（2^5，/，/）	64.70±7.06	16.33±10.09	71.70±5.49
	CELM	（2^2，2^1，/）	64.80±5.41	17.71±8.84	74.09±5.02
	SPELM	（2^2，/，0.6）	64.80±7.00	16.08±9.47	72.33±4.58
	SSPELM	（2^2，/，0.5）	**65.87±5.27**	**18.48±9.12**	**74.38±5.37**
Breast Cancer	ELM	（/，/，/）	88.89±4.32	77.84±2.89	92.72±1.00
	RELM	（2^{-1}，/，/）	89.53±3.65	75.42±4.88	91.48±1.56
	CELM	（2^3，2^3，/）	91.11±1.55	72.29±7.03	91.04±2.33
	SPELM	（2^5，/，0.5）	88.96±4.55	77.31±2.92	92.53±1.18
	SSPELM	（2^1，/，0.4）	**92.11±1.66**	**78.93±2.91**	**93.04±1.13**
Biodeg	ELM	（/，/，/）	60.66±8.99	16.62±9.59	50.11±6.91
	RELM	（2^{-3}，/，/）	**61.57±7.89**	17.75±10.08	50.97±6.84
	CELM	（2^{-1}，2^2，/）	61.47±7.85	18.48±10.09	51.17±6.80
	SPELM	（2^{-4}，/，0.6）	57.91±8.43	**19.42±3.92**	**53.15±2.91**
	SSPELM	（2^{-4}，/，0.5）	59.72±9.26	17.11±4.31	51.95±3.60
Audit_Risk	ELM	（/，/，/）	71.03±12.67	62.77±25.10	72.95±15.99
	RELM	（2^{-1}，/，/）	70.84±13.94	61.41±23.78	71.68±14.74
	CELM	（2^5，2^3，/）	70.84±13.94	61.41±23.78	71.68±14.74
	SPELM	（2^{-4}，/，0.7）	**74.00±19.28**	**65.33±26.28**	**75.39±16.50**
	SSPELM	（2^{-4}，/，0.5）	72.26±18.13	59.72±23.95	69.52±15.45
Dermatology	ELM	（/，/，/）	75.34±12.19	54.10±9.70	59.46±7.54
	RELM	（2^1，/，/）	75.71±11.88	54.25±10.18	**59.69±8.03**
	CELM	（2^5，2^2，/）	76.13±12.00	54.18±10.47	59.63±8.33
	SPELM	（2^{-2}，/，0.5）	75.71±11.88	**54.35±10.21**	**59.69±8.03**
	SSPELM	（2^0，/，0.5）	**77.47±12.76**	54.01±11.79	59.51±9.07
Data_Banknote _Authenticatio n	ELM	（/，/，/）	82.41±6.07	60.02±17.39	83.77±7.15
	RELM	（2^5，/，/）	83.14±8.24	64.62±15.06	85.69±5.89
	CELM	（2^5，2^2，/）	82.67±5.82	62.49±5.88	84.79±2.29
	SPELM	（2^5，/，0.6）	66.17±3.15	34.91±4.35	76.31±1.67
	SSPELM	（2^5，/，0.4）	**83.61±7.52**	**73.91±10.05**	**88.82±4.16**

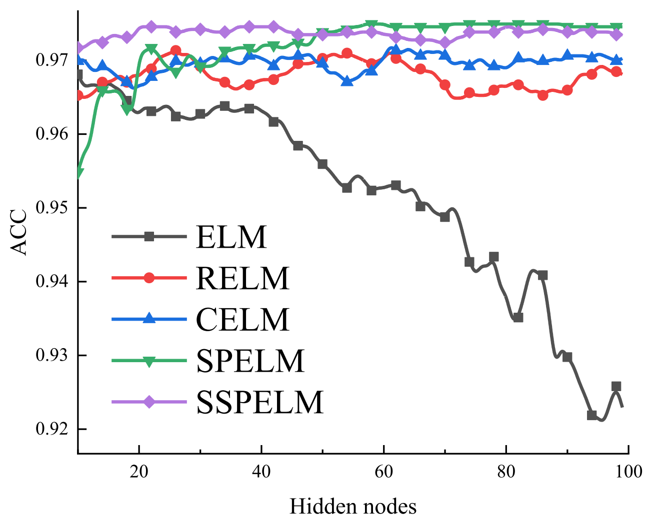

FIGURE 1. Breast cancer-sigmoid.

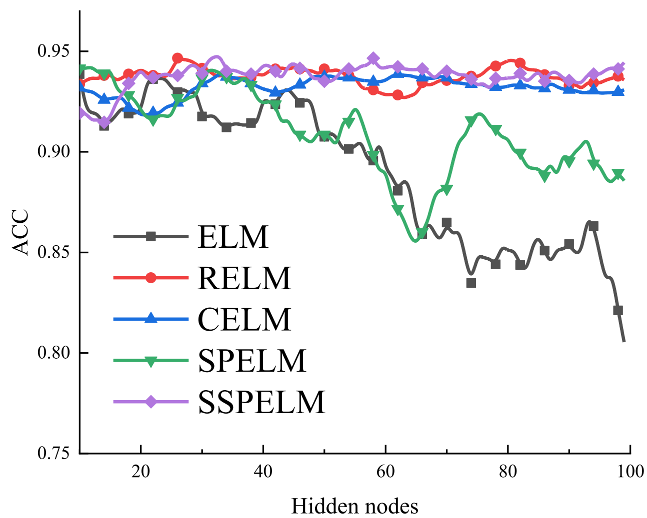

FIGURE 2. Breast cancer-sine.

By analyzing Figures 1 to 6, we can conclude that the classification accuracy of the SSPELM, RELM, and CELM algorithms improves with the increase in the number of hidden layer neurons, and ELM will exhibit abnormal performance degradation. Performance degradation also appears on the data set. On all data sets, SSPELM always performs better

TABLE 6. The experimental results.

Dataset	Method	Training time(sec)		Testing time(sec)	
		sigmoid	sine	sigmoid	sine
German	ELM	0.1552±0.009	0.1544±0.002	0.0254±0.009	0.0230±0.005
	RELM	0.1332±0.002	0.1342±0.004	0.0208±0.005	0.0268±0.006
	CELM	0.1612±0.005	0.1572±0.002	0.0334±0.025	0.0262±0.004
	SPELM	0.1474±0.006	0.1508±0.003	0.0196±0.008	0.0268±0.006
	SSPELM	0.2150±0.007	0.2346±0.024	0.0246±0.009	0.0272±0.007
Breast Cancer	ELM	0.1584±0.003	0.1620±0.005	0.0196±0.003	0.0198±0.008
	RELM	0.1420±0.005	0.1258±0.002	0.0160±0.008	0.0286±0.002
	CELM	0.1442±0.004	0.1426±0.005	0.0318±0.010	0.0238±0.005
	SPELM	0.1540±0.006	0.1516±0.006	0.0218±0.006	0.0246±0.008
	SSPELM	0.4046±0.008	0.4124±0.007	0.0266±0.010	0.0274±0.008
Biodeg	ELM	0.1784±0.005	0.1714±0.001	0.0162±0.007	0.0256±0.004
	RELM	0.1376±0.004	0.1402±0.004	0.0216±0.005	0.0186±0.003
	CELM	0.1474±0.007	0.1492±0.002	0.0276±0.014	0.0230±0.007
	SPELM	0.1628±0.008	0.1586±0.004	0.0238±0.014	0.0236±0.008
	SSPELM	0.5330±0.066	0.5962±0.010	0.0304±0.012	0.0276±0.011
Audit_Risk	ELM	0.1592±0.005	0.1616±0.003	0.0224±0.005	0.0240±0.003
	RELM	0.1362±0.006	0.1366±0.006	0.0158±0.005	0.0232±0.007
	CELM	0.1454±0.003	0.1430±0.002	0.0232±0.002	0.0266±0.007
	SPELM	0.1486±0.001	0.1536±0.005	0.0282±0.007	0.0326±0.007
	SSPELM	0.4250±0.010	0.3766±0.035	0.0394±0.009	0.0294±0.011
Dermatology	ELM	0.9412±0.006	0.9496±0.009	0.0212±0.004	0.0212±0.004
	RELM	0.7952±0.014	0.8010±0.012	0.0258±0.005	0.0258±0.005
	CELM	0.8136±0.008	0.8272±0.010	0.0289±0.005	0.0289±0.005
	SPELM	0.8410±0.007	0.8446±0.007	0.0282±0.005	0.0283±0.005
	SSPELM	1.4020±0.012	1.4530±0.023	0.0281±0.005	0.0282±0.005
Data_Banknote_Authentication	ELM	0.1598±0.004	0.1620±0.002	0.0208±0.003	0.0184±0.002
	RELM	0.1374±0.004	0.1388±0.006	0.0216±0.003	0.0246±0.003
	CELM	0.1676±0.005	0.1548±0.004	0.0238±0.002	0.0278±0.005
	SPELM	0.1662±0.002	0.1692±0.004	0.0232±0.001	0.0292±0.007
	SSPELM	0.4830±0.159	0.7464±0.012	0.0446±0.008	0.0326±0.014

TABLE 7. Aeroengine air path fault information.

Fault Code	Fault Parts	Fault Symptoms
F1	The fan blade	The fan reduced the flow by 7%
LC1	Supercharged blade	Boost stage reduced flow by 7%
HC1	Compressor blade	Compressor reduced flow rate by 4%
HT1	High pressure turbine blade	High-pressure turbines reduce flow by 6%
LT1	Low pressure turbine blade	Low pressure turbines reduce flow by 6%

than SPELM, and on most data, it can still maintain a high classification accuracy under different numbers of neurons. The performance of some data sets is comparable to CELM and RELM, but most of the data have better performance than the other four algorithms.

But in the actual application, in most cases we are unable to get data of no noise interference, discrete label data might be relatively simple to remove noise, but once input characteristics of the continuous data are polluted by noise or outliers, it is often difficult to peel off the noise data. In this paper, two new types of loss functions are introduced to study in order to develop a classification algorithm that can still perform well under the interference of noise. This plays an extremely important role in fault diagnosis of aeroengines, so we added gaussian white noise data subject to N (0,0.5) distribution to the input characteristics of the six data sets, and kept the other parameter Settings unchanged for the experiment.

TABLE 8. Aeroengine air path fault data set.

Data sets	Noise	features	training	testing
Case1	/	8	1500	1000
Case2	N (0,0.5)	8	1500	1000
Case3	N (0.1,0.5)	8	1500	1000

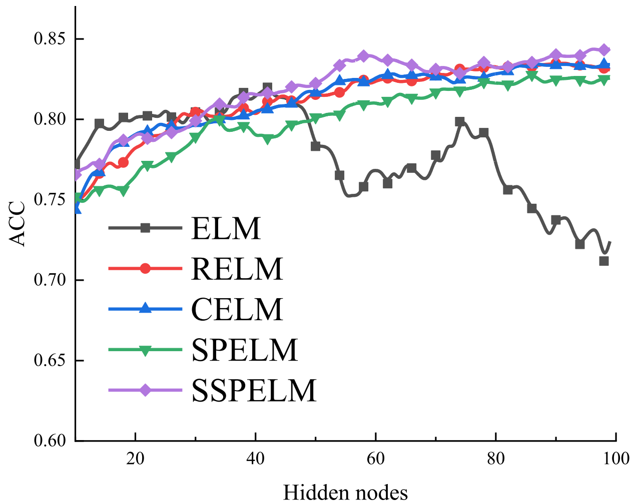

FIGURE 3. Biodeg-sigmoid.

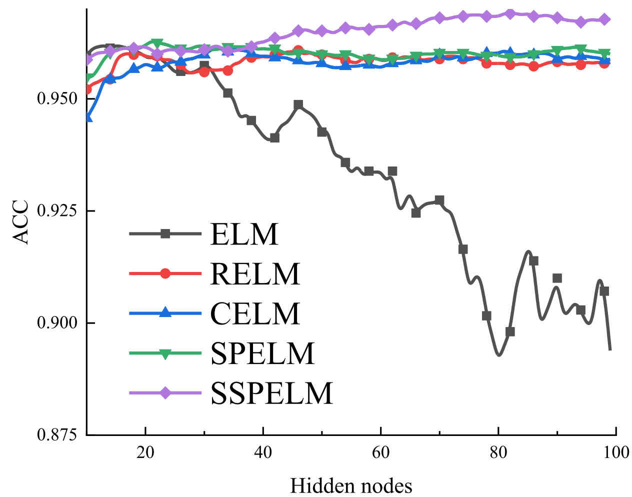

FIGURE 5. Audit_Risk-sigmoid.

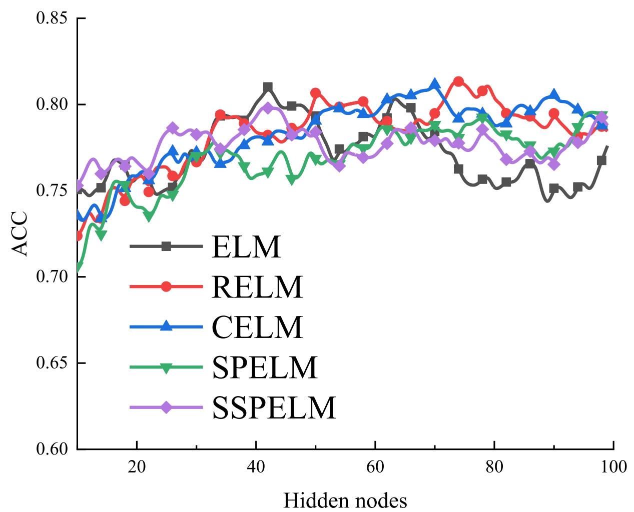

FIGURE 4. Biodeg-sine.

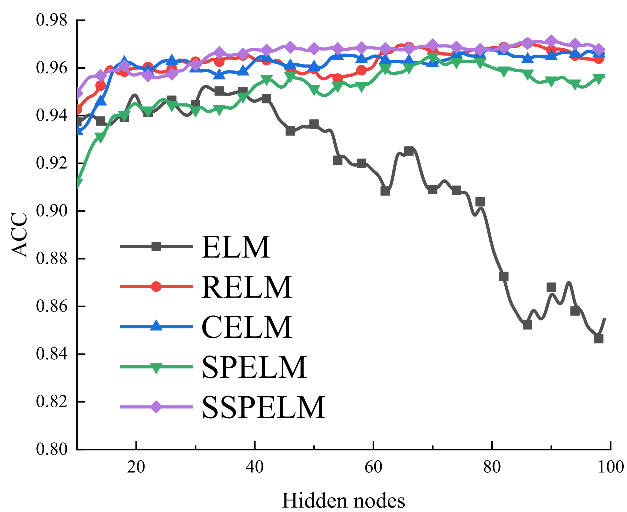

FIGURE 6. Audit_Risk-sine.

By analyzing Table 4 and Table 5, we can draw the following conclusions: Adding random white noise to the input characteristics of the training samples degrades the predictive performance of all algorithms. However, on the whole, under the two activation functions, SSPELM still maintained the optimal performance on most data sets, which proved the robustness of SSPELM; Compared with the experiment under the condition of no noise, the performance of SSPELM significantly decreased, while the performance of CELM and SPELM improved under the condition of noise, which on the one hand proved the superior robustness of C-loss function and square marble loss function. Although the introduction of soft threshold can improve the classification accuracy of the algorithm, it also affects the robustness of the loss function.

Following is a simple analysis of the computational complexity of several algorithms. ELM and RELM are non-iterative training algorithms. Once the model parameters are determined, the training time of the algorithm is very

TABLE 9. Aeroengine air path fault results.

Dataset	The type of hidden nodes	Method	(C, σ, p)	ACC (%)	MCC (%)	F_1 (%)
Case1	Sigmoid	ELM	(/, /, /)	98.95±0.11	**98.39±0.39**	98.64±0.21
		RELM	(2^5, /, /)	**99.02±0.16**	97.71±0.50	98.00±0.55
		CELM	(2^5, 2^1, /)	98.75±0.13	97.45±0.58	97.96±0.53
		SPELM	(2^5, /, 0.5)	98.88±0.13	97.54±0.52	98.03±0.50
		SSPELM	(2^5, /, 0.5)	98.71±0.11	97.32±0.60	**98.81±0.72**
	Sine	ELM	(/, /, /)	98.54±0.67	97.99±0.87	98.35±0.69
		RELM	(2^5, /, /)	98.57±0.67	97.92±0.68	98.31±0.56
		CELM	(2^5, 2^1, /)	98.65±0.69	97.81±0.68	98.21±0.56
		SPELM	(2^5, /, 0.5)	98.57±0.69	97.86±0.71	98.25±0.58
		SSPELM	(2^5, /, 0.4)	**99.10±0.43**	**98.21±0.63**	**98.57±0.71**
Case2	Sigmoid	ELM	(/, /, /)	97.78±0.81	96.64±0.99	97.11±0.80
		RELM	(2^5, /, /)	97.69±0.45	**96.88±0.68**	97.48±0.55
		CELM	(2^5, 2^1, /)	97.70±0.34	96.34±0.50	97.02±0.42
		SPELM	(2^5, /, 0.5)	97.82±0.35	96.72±0.59	97.35±0.49
		SSPELM	(2^5, /, 0.4)	**97.91±0.29**	96.00±1.09	**97.73±0.41**
	Sine	ELM	(/, /, /)	97.22±0.92	93.14±3.65	94.36±3.04
		RELM	(2^5, /, /)	97.29±0.89	93.26±3.59	94.48±2.93
		CELM	(2^5, 2^1, /)	97.27±1.08	93.96±2.93	95.03±2.41
		SPELM	(2^5, /, 0.5)	97.26±0.93	93.30±3.51	94.48±2.92
		SSPELM	(2^5, /, 0.4)	**98.40±0.75**	**95.73±1.99**	**96.72±1.53**
Case3	Sigmoid	ELM	(/, /, /)	97.01±1.70	94.71±2.20	95.35±1.72
		RELM	(2^5, /, /)	97.20±1.20	93.84±1.53	94.90±1.27
		CELM	(2^5, 2^1, /)	97.99±0.83	94.70±1.37	95.66±1.14
		SPELM	(2^5, /, 0.5)	97.30±1.08	93.62±1.34	94.74±1.12
		SSPELM	(2^5, /, 0.4)	**98.30±0.39**	**96.01±0.87**	**96.79±0.70**
	Sine	ELM	(/, /, /)	97.13±0.73	94.92±1.98	95.84±1.65
		RELM	(2^5, /, /)	97.17±0.74	95.09±1.91	95.89±1.59
		CELM	(2^5, 2^0, /)	97.60±1.81	95.87±2.15	96.63±1.81
		SPELM	(2^4, /, 0.6)	97.46±1.31	95.19±1.96	96.06±1.63
		SSPELM	(2^5, /, 0.4)	**99.07±0.36**	**97.31±1.31**	**97.78±1.06**

fast; CELM, SPELM and SSPELM are iterative training algorithms because they involve multiple variables to be optimized, and the training time must be increased compared with non-iterative training algorithms. However, since no adjustment has been made to the network structure, the real-time performance of several algorithms in the test should be better. We recorded time-related indicators such as training time and test time, as shown in Table 6. The experiment was conducted in a noiseless environment.

Based on the analysis of Table 6, we can draw the following conclusions: Although CELM, SPELM and SSPELM are iterative training algorithms that consume more training time, however, in most data sets, the test time did not change too much, showing the same speed as RELM, CELM and SPELM. This performance ensured the real-time performance of the algorithm in practical application. In the multi-classification data set, because the "one to many" strategy is adopted, the training time is the longest. But, on the test set, SSPELM still shows considerable real-time performance. The main factor influencing the training time of the algorithm is the sample size of the training data. When the sample size is large, the time consumption of SSPELM compared with other algorithms will also increase.

VI. AIR PATH FAULT DIAGNOSIS OF AEROENGINE

Through the previous numerical experiment results, it can be concluded that the SSPELM algorithm has excellent robustness and higher classification accuracy, and the number of hidden layer neurons will not affect these advantages. These characteristics make SSPELM very suitable for For the diagnosis of aero engine faults, a small number of hidden layer neurons can be used to obtain excellent diagnostic results, and the operation safety of the aero engine can be ensured when a small amount of memory is required. To further illustrate these advantages of the SSPELM algorithm in this section, we make air path fault diagnosis for the aeroengine. The selected working conditions are the standard cruise working conditions of a certain aero engine:H = 10700m, flight Mach number Ma = 0.395518, and thrust FN = 47.01kn. Five typical air path fault data were selected, as shown in Table 7, 500 of each fault data were selected. In order to verify the performance of the algorithm under noise interference, we added the gaussian white noise following N (0,0.5) and N (0.1,0.5) to the input characteristics of the original aeroengine air path fault data, and generated air path fault data under three states, each of which is shown in Table 8. Input characteristic, selection of high-pressure rotor speed $N1$, low pressure rotor speed $N2$, fan pressure π_F, pressurized pressure ratio π_{LC}, compressor pressure ratio π_{HC}, total inlet temperature of high-pressure compressor T_{25}, low pressure turbine exhaust temperature T_5, and fuel consumption W_f.

The multi-classification problems of aeroengine air path fault diagnosis, we adopt the method of "one to many" training, for each data set, we randomly according to the proportion of 60% for training samples and 40% for testing samples, each 10 times experiment data sets, the other parameters selection and set up and numerical experiment set the previous section, using K-CV to select parameters, K take 5 fold, the experimental results are shown below:

By analyzing Table 9, we can draw the following conclusions: SSPELM did not perform optimally under the Sigmoid activation function for the original noise-free aeroengine's air path data. SSPELM performed optimally under the Sine activation function. However, with the addition of gaussian white noise data to the input characteristics, the superiority of SSPELM is shown. With the addition of two kinds of noise, SSPELM still maintains a high diagnostic performance. In order to show the performance comparison of several algorithms more intuitively, we make the following figure to illustrate.

Fig.7 to Fig.9 respectively represent the evaluation indexes of ACC, MCC and F1 of the five algorithms in air path fault diagnosis. By analyzing the three graphs, it is not difficult

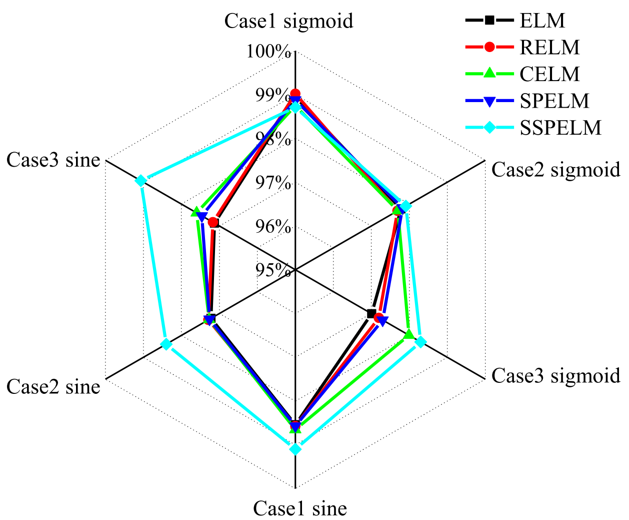

FIGURE 7. Engine air path fault diagnosis ACC.

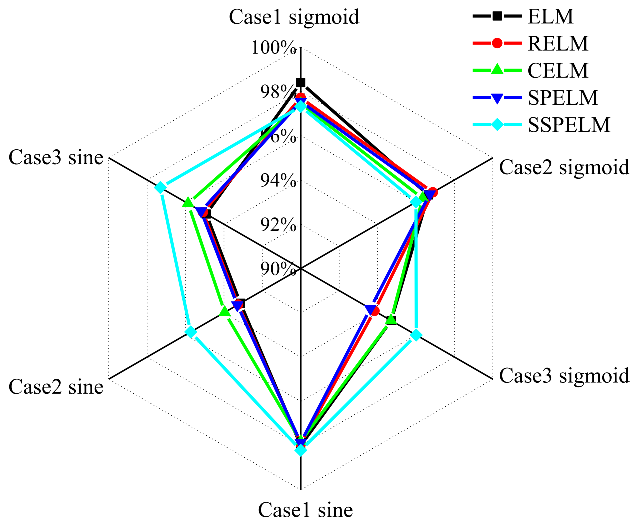

FIGURE 8. Engine air path fault diagnosis MCC.

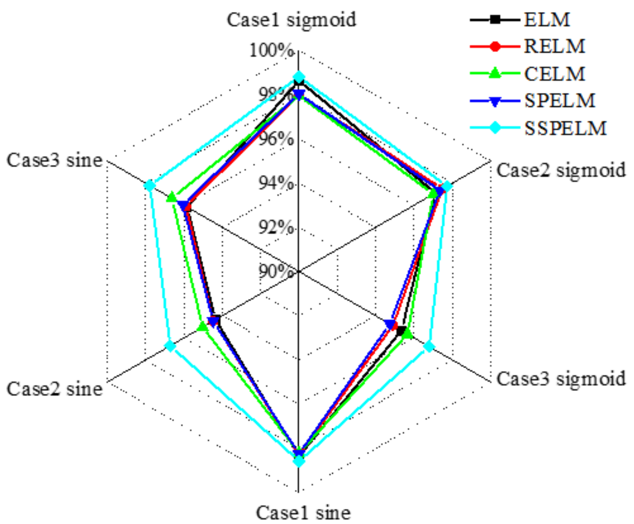

FIGURE 9. Engine air path fault diagnosis.

to draw a conclusion: SSPELM has the best performance in most cases; Under different activation functions of the same dataset, SSPELM showed greater advantages over the other four algorithms when Sine was used as the hidden layer activation function than when Sigmoid was used as the hidden layer activation function; With the increase of noise data intensity, the diagnostic accuracy of the algorithm tends to decline. However, compared with other algorithms, this trend is not obvious, which proves the excellent performance of the SSPELM algorithm in the aeroengine air path fault data set.

For aeroengine fault diagnosis, the diagnosis of the real-time performance is more important, so to test time has the certain requirement of the algorithm, the algorithm of the training time is less, the better, the longer training time is bad for us to add new data in the practical application, is unfavorable to expand into the online learning process, we here about the training time of the algorithm and the experiment of testing time, also in the above three data sets, the parameter

TABLE 10. Aeroengine air path fault results.

Dataset	The type of hidden nodes	Method	Training time(sec)	Testing time(sec)
Case1	Sigmoid	ELM	0.8270±0.005	0.1378±0.022
		RELM	0.7164±0.008	0.1284±0.018
		CELM	0.9610±0.008	0.1576±0.020
		SPELM	0.8956±0.006	0.1534±0.007
		SSPELM	4.1034±0.036	0.1968±0.057
	Sine	ELM	0.8380±0.012	0.1180±0.024
		RELM	0.7140±0.010	0.1192±0.014
		CELM	0.9490±0.007	0.1494±0.013
		SPELM	0.8868±0.006	0.1524±0.010
		SSPELM	4.1232±0.018	0.2088±0.028
Case2	Sigmoid	ELM	0.8438±0.011	0.1180±0.015
		RELM	0.7206±0.009	0.1166±0.016
		CELM	1.1301±0.007	0.1370±0.011
		SPELM	0.9008±0.014	0.1424±0.020
		SSPELM	3.7588±0.038	0.1334±0.016
	Sine	ELM	0.8336±0.007	0.1344±0.008
		RELM	0.7204±0.012	0.1166±0.020
		CELM	1.1194±0.008	0.1474±0.008
		SPELM	0.8890±0.007	0.1440±0.012
		SSPELM	4.5902±0.250	0.1762±0.073
Case3	Sigmoid	ELM	0.8426±0.011	0.1328±0.016
		RELM	0.7160±0.011	0.1244±0.019
		CELM	1.1162±0.018	0.1526±0.032
		SPELM	0.8924±0.007	0.1438±0.019
		SSPELM	3.6142±0.106	0.1794±0.018
	Sine	ELM	0.8296±0.006	0.1314±0.007
		RELM	0.7126±0.0137	0.1318±0.011
		CELM	1.1430±0.0172	0.1432±0.024
		SPELM	0.8846±0.0146	0.1496±0.026
		SSPELM	4.1876±0.2789	0.1670±0.051

Settings are unchanged, training and testing of an algorithm, 10 times results will be recorded in Table 10.

By analyzing Table 10, we can draw the following conclusions: In terms of training time, SSPELM takes longer, because SSPELM needs to consume more iterations to meet the convergence condition. The same is true for CELM, both of which require more training time than SPELM; On the test of time, SSPELM no significant degradation of performance, in some cases, even showed the optimal test time, partly to ensure the algorithm in the actual application of aeroengine fault diagnosis, because SSPELM algorithm of real-time not received too big impact, as well as maintaining high robustness and classification accuracy, is a really feasible aeroengine fault diagnosis algorithm.

VII. CONCLUSION

As a training algorithm based on the squared loss function, ELM's poor robustness and generalization limit its practical application in complex situations to a certain extent. In order to solve these two limitations, we introduce the non-convex loss function C-loss loss function and pinball loss function to obtain two robust algorithms, CELM and SPELM; in order to solve the problem of generalization, we find that The main reason for the poor generalization is that the algorithm based on the square-type loss function will force the squares of all training samples to approach 1 when performing $\{-1, +1\}$ -type label binary classification. According to the analysis of the nature of the sign discriminant function, it is not necessary to force all the training sample margins to equal 1. When the positive sample margin is greater than 0 and the negative sample is less than 0, the algorithm can obtain better performance. The introduction of this soft threshold can further improve the generalization of the algorithm. Driven by the above two solutions, we propose a new soft threshold squared pinball loss function for classification. The contributions of this paper can be summarized as follows:

1) The ELM algorithm based on the square pinball loss function and the C-loss loss function for two classifications is studied;

2) An ELM algorithm based on soft threshold pinball loss function to solve the classification problem is proposed;

3) The proposed algorithm is verified by numerical experiments on several benchmark data sets. The experimental results demonstrate the excellent robustness and generalization of the proposed algorithm;

4) The new loss function ELM is used in the aero-engine gas path fault diagnosis, and has achieved good diagnostic results.

REFERENCES

[1] Y.-P. Zhao, J.-J. Wang, X.-Y. Li, G.-J. Peng, and Z. Yang, "Extended least squares support vector machine with applications to fault diagnosis of aircraft engine," *ISA Trans.*, vol. 97, pp. 189–201, Feb. 2020.

[2] F. Lu, J. Wu, J. Huang, and X. Qiu, "Aircraft engine degradation prognostics based on logistic regression and novel OS-ELM algorithm," *Aerosp. Sci. Technol.*, vol. 84, pp. 661–671, Jan. 2019.

[3] J. Yu, "Aircraft engine health prognostics based on logistic regression with penalization regularization and state-space-based degradation framework," *Aerosp. Sci. Technol.*, vol. 68, pp. 345–361, Sep. 2017.

[4] H. Miao, B. Li, C. Sun, and J. Liu, "Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5023–5032, Sep. 2019.

[5] J. Lu, J. Huang, and F. Lu, "Distributed kernel extreme learning machines for aircraft engine failure diagnostics," *Appl. Sci.*, vol. 9, no. 8, p. 1707, Apr. 2019.

[6] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[7] L. Wang, H. Jia, and J. Li, "Training robust support vector machine with smooth ramp loss in the primal space," *Neurocomputing*, vol. 71, nos. 13–15, pp. 3020–3025, Aug. 2008.

[8] Y. Wu and Y. Liu, "Robust truncated hinge loss support vector machines," *J. Amer. Stat. Assoc.*, vol. 102, no. 479, pp. 974–983, Sep. 2007.

[9] Y. Zhao and J. Sun, "Robust support vector regression in the primal," *Neural Netw.*, vol. 21, no. 10, pp. 1548–1555, Dec. 2008.

[10] Y.-P. Zhao and J.-G. Sun, "Robust truncated support vector regression," *Expert Syst. Appl.*, vol. 37, no. 7, pp. 5126–5133, Jul. 2010.

[11] X. Yang, L. Tan, and L. He, "A robust least squares support vector machine for regression and classification with noise," *Neurocomputing*, vol. 140, pp. 41–52, Sep. 2014.

[12] F. Nie, X. Wang, and H. Huang, "Multiclass capped ℓ_p-norm SVM for robust classifications," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 2415–2421.

[13] A. Singh, R. Pokharel, and J. Principe, "The C-loss function for pattern classification," *Pattern Recognit.*, vol. 47, no. 1, pp. 441–453, Jan. 2014.

[14] G. Xu, B.-G. Hu, and J. C. Principe, "Robust C-loss kernel classifiers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 510–522, Mar. 2018.

[15] B. Chen, X. Wang, Y. Li, and J. C. Principe, "Maximum correntropy criterion with variable center," 2019, *arXiv:1904.06501*. [Online]. Available: http://arxiv.org/abs/1904.06501

[16] Y.-P. Zhao, J.-F. Tan, J.-J. Wang, and Z. Yang, "C-loss based extreme learning machine for estimating power of small-scale turbojet engine," *Aerosp. Sci. Technol.*, vol. 89, pp. 407–419, Jun. 2019.

[17] I. Steinwart and A. Christmann, "Estimating conditional quantiles with the help of the pinball loss," *Bernoulli*, vol. 17, no. 1, pp. 211–225, Feb. 2011.

[18] X. Huang, L. Shi, and J. A. K. Suykens, "Support vector machine classifier with pinball loss," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 984–997, May 2014.

[19] X. Huang, L. Shi, and J. A. K. Suykens, "Asymmetric least squares support vector machine classifiers," *Comput. Statist. Data Anal.*, vol. 70, pp. 395–405, Feb. 2014.

[20] M. Farooq and I. Steinwart, "An SVM-like approach for expectile regression," *Comput. Statist. Data Anal.*, vol. 109, pp. 159–181, May 2017.

[21] Y. Xu, Z. Yang, and X. Pan, "A novel twin support-vector machine with pinball loss," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 2, pp. 359–370, Feb. 2017.

[22] R. Gong, C. Wu, M. Chu, and H. Wang, "Twin pinball loss support vector hyper-sphere classifier for pattern recognition," in *Proc. Chin. Control Decis. Conf. (CCDC)*, May 2016, pp. 6551–6556.

[23] Y. Xu, Q. Wang, X. Pang, and Y. Tian, "Maximum margin of twin spheres machine with pinball loss for imbalanced data classification," *Int. J. Speech Technol.*, vol. 48, no. 1, pp. 23–34, Jan. 2018.

[24] Y. Xu, Z. Yang, Y. Zhang, X. Pan, and L. Wang, "A maximum margin and minimum volume hyper-spheres machine with pinball loss for imbalanced data classification," *Knowl.-Based Syst.*, vol. 95, pp. 75–85, Mar. 2016.

[25] W. Zhu, Y. Song, and Y. Xiao, "A new support vector machine plus with pinball loss," *J. Classification*, vol. 35, no. 1, pp. 52–70, Apr. 2018.

[26] L. Yang and H. Dong, "Support vector machine with truncated pinball loss and its application in pattern recognition," *Chemometric Intell. Lab. Syst.*, vol. 177, pp. 89–99, Jun. 2018.

[27] X. Shen, L. Niu, Z. Qi, and Y. Tian, "Support vector machine classifier with truncated pinball loss," *Pattern Recognit.*, vol. 68, pp. 199–210, Aug. 2017.

[28] K. Wang and X. Ding, "Pinball loss based extreme learning machines," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 569, Aug. 2019, Art. no. 052061.

[29] Y.-P. Zhao, G. Huang, Q.-K. Hu, J.-F. Tan, J.-J. Wang, and Z. Yang, "Soft extreme learning machine for fault detection of aircraft engine," *Aerosp. Sci. Technol.*, vol. 91, pp. 70–81, Aug. 2019.

[30] W. Deng, Q. Zheng, and L. Chen, "Regularized extreme learning machine," in *Proc. IEEE Symp. Comput. Intell. Data Mining*, Mar. 2009, pp. 389–395.

[31] M. Allain, J. Idier, and Y. Goussard, "On global and local convergence of half-quadratic algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1130–1142, May 2006.

[32] S.-H. Yang and B.-G. Hu, "A stagewise least square loss function for classification," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2008, pp. 120–131.

[33] A. Frank and A. Asuncion. (2010). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml?

[34] Y. Zhao and K. Wang, "Fast cross validation for regularized extreme learning machine," *J. Syst. Eng. Electron.*, vol. 25, no. 5, pp. 895–900, Oct. 2014.

YUYUAN CAO received the B.S. degree from the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China, and the M.S. degree in aviation safety management from the French National University of Civil Aviation and the French University of Aviation, in 2009. He is currently a Researcher with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics. His current research interests include aeroengine fault diagnosis and machine learning.

BOWEN ZHANG received the B.S. degree in aircraft airworthiness technology from the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2014, where he is currently pursuing the master's degree. His current research interests include aeroengine fault diagnosis and machine learning.

HUAWEI WANG received the Ph.D. degree from the National University of Defense Technology. She is currently a Professor with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics. Her current research interests include carrying vehicle application engineering: aircraft reliability engineering and civil aircraft maintenance engineering; safety science and technology: civil aviation safety analysis, assessment and control, and civil aviation safety management; and aircraft airworthiness technology: aircraft safety analysis technology and airworthiness management.

YU BAI received the B.S. degree in transportation engineering from the College of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China, in 2013. He is currently pursuing the master's degree with the Nanjing University of Aeronautics and Astronautics, Nanjing, China. His current research interests include aeroengine fault diagnosis and machine learning.