Neural Processing Letters (2024) 56:116
https://doi.org/10.1007/s11063-024-11492-5

Intuitionistic Fuzzy Extreme Learning Machine with the Truncated Pinball Loss

Qingyun Gao[1] · Qing Ai[1] · Wenhui Wang[2]

Accepted: 17 October 2023

Abstract
Fuzzy extreme learning machine (FELM) is an effective algorithm for dealing with classification problems with noises, which uses a membership function to effectively suppress noise in data. However, FELM has the following drawbacks: (a) The membership degree of samples in FELM is constructed by considering only the distance between the samples and the class center, not the local information of samples. It is easy to mistake some boundary samples for noises. (b) FELM uses the least squares loss function, which leads to sensitivity to feature noise and instability to re-sampling. To address the above drawbacks, we propose an intuitionistic fuzzy extreme learning machine with the truncated pinball loss (TPin-IFELM). Firstly, we use the K-nearest neighbor (KNN) method to obtain local information of the samples and then construct membership and non-membership degrees for each sample in the random mapping feature space based on valuable local information. Secondly, we calculate the score value of samples based on the membership and non-membership degrees, which can effectively identify whether the boundary samples are noises or not. Thirdly, in order to maintain the sparsity and robustness of the model, and enhance the stability of the resampling of the model, we introduce the truncated pinball loss function into the model. Finally, in order to solve more efficiently, we employ the concave-convex procedure (CCCP) to solve TPin-IFELM. Extensive comparative experiments are conducted on the benchmark datasets to verify the superior performance of TPin-IFELM.

Keywords Extreme learning machine · Intuitionistic fuzzy sets · Truncated pinball loss · Noise insensitivity

✉ Qing Ai
lyaiqing@126.com

Qingyun Gao
ustlgqy@163.com

Wenhui Wang
angleboy@foxmail.com

[1] School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China

[2] College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

Published online: 20 March 2024

1 Introduction

Extreme learning machine (ELM) [1] is a simple and efficient single-hidden layer feedforward neural network (SLFN) that is much faster than other feedforward neural networks. The input weights and hidden layer biases of ELM are randomly generated, and the network output weights are obtained by using Moore-Penrose generalized inverse. The objective of ELM is to achieve the minimal norm of output weights and the minimum training error. In recent years, to improve the classification performance of ELM, some improvements have been proposed [2–12]. Due to their excellent classification performance, these algorithms have been applied to a wide range of fields [13–23].

In practical classification problems, there is a large amount of noise in the data. Noise can interfere with the construction of classifiers and reduce the classification performance of algorithms. However, traditional ELMs cannot effectively suppress the negative impact of noise. To enhance the noise robustness of ELM, Ren et al. [24] proposed the correntropy-based hinge loss robust extreme learning machine (CHELM). Wang et al. [25] proposed the extreme learning machine with the homotopy loss (l_1-HELM), which introduces the homotopy loss into ELM. To enhance the noise robustness and re-sampling stability, Ren et al. [26] proposed the extreme learning machine with the pinball loss function (PELM), which maximizes the quantile distance between two classes of samples. To further achieve robustness and sparsity, Shen et al. [27] introduced ε-insensitive zone pinball loss into the ELM. In practical classification problems, there are a large number of redundant or irrelevant features in the data. These algorithms are negatively affected by these features, which reduces their performance. To reduce redundant or irrelevant features, Huang et al. [28] propose a method that employs a new fuzzy β neighborhood-related discernibility measure and the fuzzy β covering (FBC) decision tables. To enhance the robustness of FBC in feature learning, Huang et al. [29] propose a noise-tolerant fuzzy-β-covering-based multigranulation rough set model. To deal with noisy data, VPDI uses noise-tolerant discrimination indexes and a heuristic feature selection algorithm to reduce redundant or irrelevant features [30]. Although these algorithms based on the FBC feature selection can eliminate redundant or irrelevant features, they do not consider the importance of samples in the classification process. In recent years, determining the importance of samples has become a research hotspot. Inspired by FSVM [31], Zhang et al. [32] proposed the fuzzy extreme learning machine (FELM). FELM employs a membership degree to each training sample, which can reduce the influence of outliers and noise.

FELM is an effective algorithm for dealing with classification problems with noises. However, it has two drawbacks: (1) FELM only considers the membership degree of samples but not the non-membership degree of samples, which can easily mistake the boundary samples for noise. (2) FELM uses the least square loss function, which leads to sensitivity to the feature noise and instability to re-sampling. To address the above drawbacks, we propose an improved ELM model combining IFSs and truncated pinball loss function, called an intuitionistic fuzzy extreme learning machine with the truncated pinball loss (TPin-IFELM). First, TPin-IFELM constructs the membership and non-membership degrees based on the local information of samples obtained by using the KNN method. The membership degree is calculated by the distance between the sample and the class center, and the non-membership degree is calculated by the correlation between all heterogeneous samples and all samples in its neighborhood. Further, we obtain the score value of the sample according to the membership and non-membership degrees. The score can be used to effectively identify whether boundary samples are noises. Finally, in order to further reduce the negative effects of noises,

we introduce the truncated pinball loss function [33, 34] into TPin-IFELM, which makes TPin-IFELM more robust and sparse. Since TPin-IFELM is a non-convex problem, we use the CCCP [35, 36] to solve it. A large number of experimental results show that the proposed TPin-IFELM is superior to some state-of-the-art comparison algorithms in dealing with classification problems with noises.

The rest of this paper is organized as follows. In Sect. 2, we briefly review ELM and its loss functions, and FELM and its improvement. In Sect. 3, we discuss the optimization model for the linear and nonlinear TPin-IFELM in detail. In Sect. 4, we investigate some properties of TPin-IFELM. In Sect. 5, the TPin-IFELM is evaluated via a series of experiments. Section 6 summarizes this paper and puts forward the future research direction.

2 Related Works

2.1 Notations

We define the binary dataset $D = \{(x_i, t_i) \mid 1 \le i \le N\}$, where $t_i = \{+1, -1\}$. Let $\mathcal{X}^+ = \{x_i \mid (x_i, t_i) \in D, t_i = +1\}$ denote the positive samples, $\mathcal{X}^- = \{x_j \mid (x_j, t_j) \in D, t_j = -1\}$ denote the negative samples, $N^+ = |\mathcal{X}^+|$, $N^- = |\mathcal{X}^-|$, $\mathcal{X} = \mathcal{X}^+ \cup \mathcal{X}^-$, and $N = N^+ + N^-$.

2.2 ELM and its Loss Functions

ELM is an effective single-layer feedforward neural network [2, 37]. First, the input weights and hidden layer biases are randomly assigned, then the hidden layer matrix H is obtained by the activation function $G(\cdot)$, and finally the output weights β are obtained by solving the generalized inverse.

The output of ELM is defined as follows:

$$f(x) = \sum_{i=1}^{L} G({\theta_i}^{\mathrm{T}} x + \vartheta_i)\beta_i = h(x) \cdot \beta, \tag{1}$$

where $\theta_i = [\theta_{i1}, \ldots, \theta_{in}]^{\mathrm{T}} \in \Re^n$ and $\vartheta_i \in \Re$ are the input weight vector and bias of the corresponding hidden node, respectively, $h(x) = \left[G({\theta_1}^{\mathrm{T}} x + \vartheta_1), \ldots, G({\theta_L}^{\mathrm{T}} x + \vartheta_L)\right]^{\mathrm{T}} \in \Re^L$ is the random feature mapping output of the hidden layer, and $\beta = [\beta_1, \beta_2, \ldots, \beta_L]^{\mathrm{T}} \in \Re^L$.

β can be solved by solving

$$min \|\beta\| \text{ and } min \sum_{i=1}^{N} \|\beta \cdot h(x_i) - t_i\|. \tag{2}$$

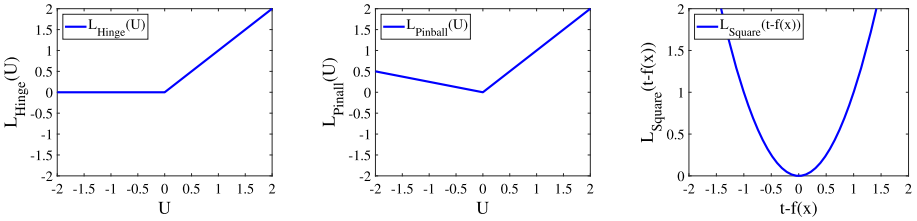
The optimal solution of (2) can be calculated by

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

where $H = [h(x_1), \ldots, h(x_N)]^{\mathrm{T}} \in \Re^{N \times L}$ and $T = [t_1, t_2, \ldots, t_N]^{\mathrm{T}} \in \Re^N$ is the output vector. $H^{\dagger}$ is the Moore-Penrose generalized inverse of matrix H.

The decision function of ELM is

$$f(x) = sign\left(h(x)^{\mathrm{T}} \hat{\beta}\right). \tag{4}$$



(a) Hinge loss (b) Pinball loss (c) Least square loss

Fig. 1 Illustrations of loss functions

In order to improve the classification performance of ELM, Huang et al. [3] proposed the optimization method-based ELM (OELM), which introduces the hinge loss function into ELM. To speed up the solution, Huang et al.[2] proposed the regular ELM (RELM), which introduces the least squares loss function into ELM. However, OELM and RELM are sensitive to noise. To enhance the noise robustness of ELM, Ren et al. [26] proposed the extreme learning machine with the pinball loss function (PELM), which maximizes the quantile distance between two classes of samples.

For convenience, we unify the optimization problems of these algorithms as follows:

$$min_{\beta}\ \frac{1}{2}\|\beta\|^2 + c\sum_{i=1}^{N} L\,(\mathrm{U}), \tag{5}$$

where $L\,(\mathrm{U})$ is the loss function. When $L\,(\mathrm{U})$ is the hinge loss or pinball loss function, $\mathrm{U} = 1 - t_i h\,(x_i)\cdot\beta$ (see Fig. 1a, b); When $L\,(\mathrm{U})$ is the least squares loss function, $\mathrm{U} = t - h\,(x_i)\cdot\beta$ (see Fig. 1c).

2.3 FELM and Its Improvements

FELM [32] employs a membership degree to each training sample to reduce the influence of outliers and noise. The optimization problem of FELM can be formulated as follows:

$$\begin{aligned} &min_{\beta,\xi_i}\ \frac{1}{2}\|\beta\|^2 + \frac{c}{2}\sum_{i=1}^{N} s_i\|\xi_i\|^2 \\ &s.t.\ \ h\,(x_i)\cdot\beta = t_i - \xi_i,\ i = 1,\ldots,N, \end{aligned} \tag{6}$$

where ξ_i is the training error, c is the penalty parameter, $s_i = \begin{cases} 1 - \frac{\|h(x_i) - \widetilde{\mathcal{C}}^+\|}{r^+ + \delta}, & t_i = +1 \\ 1 - \frac{\|h(x_i) - \widetilde{\mathcal{C}}^-\|}{r^- + \delta}, & t_i = -1 \end{cases}$ is the membership degree of x_i in the random mapping feature space, $\widetilde{\mathcal{C}}^+ = \frac{1}{N^+}\sum_{x_i\in\mathcal{X}^+} h\,(x_i)$ and $\widetilde{\mathcal{C}}^- = \frac{1}{N^-}\sum_{x_i\in\mathcal{X}^-} h\,(x_i)$ are the centers of the positive class and negative class, respectively, and $r^+ = \max_{x_i\in\mathcal{X}^+}(\|h\,(x_i) - \widetilde{\mathcal{C}}^+\|)$ and $r^- = \max_{x_i\in\mathcal{X}^-}(\|h\,(x_i) - \widetilde{\mathcal{C}}^-\|)$ are the radius of the positive class and negative class, respectively.

However, FELM only considers the membership degree of samples, which can easily mistake some boundary samples for noise. To identify the noise in the support vectors, Rezvani et al. proposed intuitionistic fuzzy twin support vector machines (IFTSVM), which use the intuitionistic fuzzy sets (IFSs) to construct the score values of samples [38]. In order to

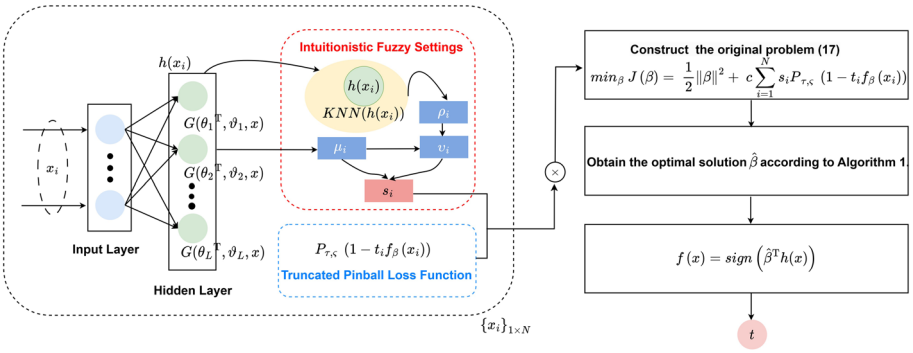


Fig. 2 The framework diagram of TPin-IFELM

enhance the robustness and re-sampling stability of IFTSVM, Liang et al. proposed an intuitionistic fuzzy twin support vector machines with the ε-insensitive pinball loss (PIFTSVM) [39]. It defines the score function named SFA:

$$s(x) = \sqrt{\frac{\mu(x)^2 + (1 - \nu(x))^2}{2}}, \tag{7}$$

where $\mu(x)$ is the membership function, and $\nu(x)$ is the non-membership function. Laxmi et al. proposed multi-category intuitionistic fuzzy twin support vector machines to solve the multi-class classification problems [40]. In order to effectively solve the problem of class imbalance, Rezvani et al. proposed class balance learning using fuzzy ART and intelligent fuzzy twin support vector machines.

3 Intuitionistic Fuzzy Extreme Learning Machines with the Truncated Pinball Loss

In this section, we propose TPin-IFELM to address the drawbacks of FELM. The algorithm framework of TPin-IFELM is shown in Fig. 2.

3.1 Intuitionistic Fuzzy Settings

FELM easily mistakes some boundary samples for noise due to its only using membership degree. To address this issue, in this subsection, We employ IFS for each sample to reduce the negative impact of noise.

Define an intuitionistic fuzzy set $\bar{A} = \left\{\left(x, \mu_{\bar{A}}(x), \nu_{\bar{A}}(x)\right) \mid x \in \mathcal{X}\right\}$, where $\mu_{\bar{A}}: \mathcal{X} \to [0, 1]$ is the membership degree of x in $\mathcal{X}$, $\nu_{\bar{A}}: \mathcal{X} \to [0, 1]$ is the non-membership degree of x in $\mathcal{X}$, and $0 \leq \mu_{\bar{A}}(x) + \nu_{\bar{A}}(x) \leq 1$. We illustrate the acquisition of membership and non-membership degrees through the following examples.

3.1.1 Intuitionistic Fuzzy Membership Degree

In the random mapping feature space, the membership degree of samples is determined by the distance between samples and the class center, i.e.,

$$\mu_i = \begin{cases} 1 - \frac{\|h(x_i) - \widetilde{\mathcal{C}}^+\|}{r^+ + \varrho}, & t_i = +1 \\ 1 - \frac{\|h(x_i) - \widetilde{\mathcal{C}}^-\|}{r^- + \varrho}, & t_i = -1 \end{cases}, \tag{8}$$

where $1 \leq i \leq N$, $\varrho > 0$ is an adjustable parameter in the random mapping feature space.

Example 1 Let $h(x_*) = (0.91, 0.27, 0.21, 0.22, 0.23)$, $t_* = +1$, $\widetilde{\mathcal{C}}^+ = (0.80, 0.52, 0.40, 0.57, 0.43)$ is the center of the positive samples, and $r^+ = 0.87$ is the radius of positive samples. According to Eq. (8), $\mu_* = 1 - \frac{0.5227}{0.87 + 10^{-7}} = 0.3992$.

3.1.2 Intuitionistic Fuzzy Non-membership Degree

We can effectively capture the correlation between x_i and all heterogeneous samples in its neighborhood by using the KNN method, i.e.,

$$\rho(x_i) = \frac{|\{x_j \mid h(x_j) \in KNN(h(x_i)), t_j \neq t_i\}|}{K}, \tag{9}$$

where $KNN(h(x_i))$ is used to represent the K nearest neighbors of x_i in the random mapping feature space.

The non-membership degree υ_i is defined as:

$$\upsilon_i = (1 - \mu_i)\, \rho(x_i), \tag{10}$$

and $0 \leq \mu_i + \upsilon_i \leq 1$.

Example 2 Let $K = 5$ and $\rho(x_i) = \frac{4}{5}$. According to Eq. (10), $\upsilon_* = (1 - 0.3992) \times \frac{4}{5} = 0.4806$.

3.1.3 The Score Function

We construct an IFS $\breve{S} = \{(x_1, t_1, \mu_1, \nu_1), (x_2, t_2, \mu_2, \nu_2), \ldots, (x_N, t_N, \mu_N, \nu_N)\}$. According to $\breve{S}$, we construct the score value (SV) as follows:

$$s_i = \begin{cases} \mu_i, & \nu_i = 0 \\ 0, & \mu_i \leq \nu_i \\ \frac{1-\nu_i}{2-\mu_i-\nu_i}, & \mu_i > \nu_i \text{ and } \nu_i \neq 0 \end{cases}, 1 \leq i \leq N, \tag{11}$$

where $s_i = \mu_i$ indicates that x_i is a correctly classified sample; $s_i = 0$ indicates that x_i is the noise; $s_i = \frac{1-\nu_i}{2-\mu_i-\nu_i}$ indicates that x_i is the support vector of the corresponding class, not the noise.

3.2 Linear Case

Unlike FELM [32] which uses the least squares loss function, TPin-IFELM employs the truncated pinball loss function, which not only makes the model robust to the noises but also preserves the sparsity. The truncated pinball loss function (see Fig. 3) is as follows:

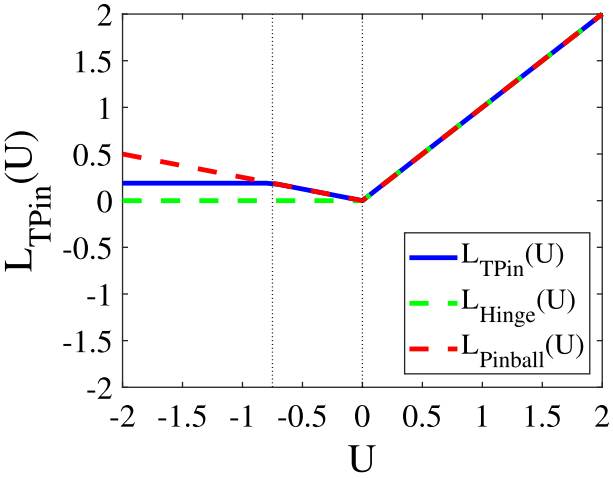


Fig. 3 The truncated pinball loss

$$P_{\tau,\varsigma}(x,t,f(x)) = \begin{cases} \tau\varsigma, & \mathrm{U} \le -\varsigma \\ -\tau\mathrm{U}, & -\varsigma < \mathrm{U} < 0 \\ \mathrm{U}, & \mathrm{U} \ge 0 \end{cases}, \tag{12}$$

where $0 \le \tau \le 1$, and $\varsigma > 0$ is the preset value, and t is the label of x.

As shown in Fig. 3, the truncated pinball loss function takes into account the advantages of the hinge loss function and pinball loss function, so it has noise robustness and sparsity.

Equation (12) can be decomposed as follows:

$$P_{\tau,\varsigma}(x,t,f(x)) = H_{1+\tau}(1-tf(x)) - (H_\tau(1-tf(x)+\varsigma)-\tau\varsigma), \tag{13}$$

where $H_{1+\tau}(1-tf(x)) = (1+\tau)\,max(0,1-tf(x))$ and $H_\tau(1-tf(x)+\varsigma) = \tau max(0,1-tf(x)+\varsigma)$.

We replace the least squares loss function in Eq. (6) with the truncated pinball loss and employ the score value for each sample as follows:

$$min_\beta\ J(\beta) = \frac{1}{2}\|\beta\|^2 + c\sum_{i=1}^{N} s_i P_{\tau,\varsigma}\left(1-t_i f_\beta(x_i)\right), \tag{14}$$

where c is the penalty parameter.

The gradient $\nabla_\beta(J(\beta))$ of $J(\beta)$ is as follows:

$$\nabla_\beta(J(\beta)) = \beta - c\sum_{i=1}^{N} s_i t_i h(x_i)\,\partial_\beta P_{\tau,\varsigma}\left(1-t_i f_\beta(x_i)\right). \tag{15}$$

It can be proved that the minimum of (14) with respect to β should satisfy the following condition

$$\beta = c\sum_{i=1}^{N} s_i t_i h(x_i)\,\partial_\beta P_{\tau,\varsigma}\left(1-t_i f_\beta(x_i)\right). \tag{16}$$

The function $J(\beta)$ in (14) can be decomposed into the sum of the convex function $J_{vex}(\beta)$ and the concave function $J_{cav}(\beta)$, i.e.,

$$J(\beta) = \underbrace{\frac{1}{2}\|\beta\|^2 + c\sum_{i=1}^{N} s_i H_{1+\tau}\left(1 - t_i f_\beta(x_i)\right)}_{J_{vex}(\beta)} \underbrace{-c\sum_{i=1}^{N} s_i H_\tau\left(1 - t_i f_\beta(x_i) + \varsigma\right) + c\tau\varsigma\sum_{i=1}^{N} s_i}_{J_{cav}(\beta)}. \tag{17}$$

Obviously, (17) is a non-differentiable non-convex optimization problem, which can be solved by the CCCP. The detailed procedure of the CCCP is shown in Algorithm 1.

Algorithm 1 CCCP for optimization problem (17)

Input:
$\beta^{(0)}$: the initial point
σ : the preset threshold
Output:
$\hat{\beta} = \beta^{(k)}$: the optimal solution
procedure :

1. Let $k = 0$ and the loss $\varepsilon = +\infty$.
2. **while** $(\varepsilon > \sigma)$
   (a) $k = k + 1$.
   (b) $\beta^{(k)} = \arg min_\beta J_{vex}(\beta) + \nabla_\beta\left(J_{cav}\left(\beta^{(k-1)}\right)\right)^{\mathrm{T}}\beta$.
   (c) $\varepsilon = \left\|\beta^{(k)} - \beta^{(k-1)}\right\|$.
3. **end while**
4. **return** $\beta^{(k)}$

end procedure

Using the CCCP to solve (17), the subproblem of the kth iteration can be expressed as:

$$\begin{aligned} &min_\beta\ J_{vex}(\beta) + \nabla_\beta\left(J_{cav}\left(\beta^{(k-1)}\right)\right)^{\mathrm{T}}\beta \\ &= \frac{1}{2}\|\beta\|^2 + c\sum_{i=1}^{N} s_i H_{1+\tau}\left(1 - t_i f_\beta(x_i)\right) + \sum_{i=1}^{N}\delta_i^{k-1} t_i f_\beta(x_i), \end{aligned} \tag{18}$$

where

$$\delta_i^{k-1} = \begin{cases} cs_i\tau, & t_i f_{\beta^{k-1}}(x_i) = t_i\left(h(x_i)\cdot\beta^{k-1}\right) < \varsigma + 1 \\ 0, & \text{otherwise} \end{cases}. \tag{19}$$

By introducing the slack variables $\xi = [\xi_1, \ldots, \xi_N]^{\mathrm{T}}$, (18) is equivalent to the following form:

$$\begin{aligned} &min_{\beta,\xi_i}\ \frac{1}{2}\|\beta\|^2 + c\sum_{i=1}^{N} s_i\xi_i + \sum_{i=1}^{N}\delta_i^{k-1} t_i f_\beta(x_i) \\ &s.t.\ t_i f_\beta(x_i) \geq 1 - \frac{1}{1+\tau}\xi_i,\ \xi_i \geq 0,\ i = 1\ldots N. \end{aligned} \tag{20}$$

According to the Lagrange method, we can obtain the following dual problem. The detailed solution process is shown in Appendix A.

$$\begin{aligned} &min_{\alpha}\ \frac{1}{2}(\alpha^{\mathrm{T}}-\delta^{\mathrm{T}})Q(\alpha-\delta)-e^{\mathrm{T}}\alpha \\ &s.t.\ \ 0\leq\ \alpha\leq(1+\tau)\,cS, \end{aligned} \tag{21}$$

where $Q = THH^{\mathrm{T}}T$.

Set $\lambda = \alpha - \delta$, and the lower and upper bounds of the box constraint are defined as $\mathfrak{L} = -\delta \in \Re^N$ and $\mathfrak{U} = (1+\tau)\,cS - \delta \in \Re^N$. Then, (21) is equivalent to

$$\begin{aligned} &min_{\lambda}\ \frac{1}{2}\lambda Q\lambda-e^{\mathrm{T}}\alpha \\ &s.t.\ \ \mathfrak{L}\leq\ \lambda\leq\mathfrak{U}. \end{aligned} \tag{22}$$

The label t of the unknown sample x is determined by the following decision function.

$$f\,(x) = sign\left(\lambda^{\mathrm{T}}THh(x)\right). \tag{23}$$

The complete process of linear TPin-IFELM is shown in Algorithm 2.

Algorithm 2 The complete process of linear TPin-IFELM

Input:

$\{(x_i, y_i) \mid 1 \leq i \leq N\}$: the training set
L: the number of hidden layer nodes
c, ς, τ and K: the parameters
x: an unknown sample

Output:

t : the label of x

procedure :

(1) Randomly assign the input weights and hidden layer biases.
(2) Obtain the hidden layer matrix H by the activation function $G(\cdot)$.
(3) Obtain the score value s_i of each training sample in the random mapping feature space according to Eq. (11).
(4) Obtain the optimal solution $\hat{\beta}$ according to Algorithm 1.
(5) Determine the label t of the unknown sample x according to (23).

end procedure

3.3 Nonlinear Case

In the ELM kernel space [2, 3, 41], the membership degree of the sample is defined by

$$\mu_i^{\Phi} = \begin{cases} 1-\dfrac{\sqrt{\mathcal{K}_{ELM}(x_i,x_i)-\frac{2}{N^+}\sum_{x_j\in\mathcal{X}^+}\mathcal{K}_{ELM}(x_i,x_j)+\frac{1}{N^{+2}}\sum_{x_i\in\mathcal{X}^+}\sum_{x_j\in\mathcal{X}^+}\mathcal{K}_{ELM}(x_i,x_j)}}{r^+ + \varrho}, & t_i = +1 \\ 1-\dfrac{\sqrt{\mathcal{K}_{ELM}(x_i,x_i)-\frac{2}{N^-}\sum_{x_j\in\mathcal{X}^-}\mathcal{K}_{ELM}(x_i,x_j)+\frac{1}{N^{-2}}\sum_{x_i\in\mathcal{X}^-}\sum_{x_j\in\mathcal{X}^-}\mathcal{K}_{ELM}(x_i,x_j)}}{r^- + \varrho}, & t_i = -1 \end{cases}, \tag{24}$$

and the non-membership degree of the sample is defined as

$$\nu_i^{\Phi} = (1-\ \mu_i^{\Phi})\rho\,(x_i)\,, \tag{25}$$

where $\mathcal{K}_{ELM}(x_i, x_i) = h(x_i) \cdot h(x_i)$, $\mathcal{K}_{ELM}(x_i, x_j) = h(x_i) \cdot h(x_j)$, $1 \le i \le N$,

$$r^+ = \max_{x_i \in \mathcal{X}^+} \left(\sqrt{\mathcal{K}_{ELM}(x_i, x_i) - \frac{2}{N^+} \sum_{x_j \in \mathcal{X}^+} \mathcal{K}_{ELM}(x_i, x_j) + \frac{1}{N^{+2}} \sum_{x_i \in \mathcal{X}^+} \sum_{x_j \in \mathcal{X}^+} \mathcal{K}_{ELM}(x_i, x_j)} \right),$$

and

$$r^- = \max_{x_i \in \mathcal{X}^-} \left(\sqrt{\mathcal{K}_{ELM}(x_i, x_i) - \frac{2}{N^-} \sum_{x_j \in \mathcal{X}^-} \mathcal{K}_{ELM}(x_i, x_j) + \frac{1}{N^{-2}} \sum_{x_i \in \mathcal{X}^-} \sum_{x_j \in \mathcal{X}^-} \mathcal{K}_{ELM}(x_i, x_j)} \right).$$

According to Eq. (24) and Eq. (25), the score function is defined as follows:

$$s_i^\Phi = \begin{cases} \mu_i^\Phi, & \nu_i^\Phi = 0 \\ 0, & \mu_i^\Phi \le \nu_i^\Phi \\ \frac{1-\nu_i^\Phi}{2-\mu_i^\Phi-\nu_i^\Phi}, & \mu_i^\Phi > \nu_i^\Phi \text{ and } \nu_i^\Phi \ne 0 \end{cases}. \tag{26}$$

The original problem of nonlinear TPin-IFELM can be expressed as

$$min_\varpi \; J(\varpi) = \frac{1}{2} \|\varpi\|^2 + C \sum_{i=1}^{N} s_i^\Phi P_{\tau,\varsigma} \left(1 - t_i \varpi^{\mathrm{T}} h(x_i)\right). \tag{27}$$

Similar to linear TPin-IFELM, (27) can be solved by CCCP. In the kth iteration, the subproblem of (27) can be expressed as

$$\begin{aligned} min_{\varpi,\xi} \; & \frac{1}{2} \|\varpi\|^2 + c \sum_{i=1}^{N} s_i^\Phi \xi_i + \sum_{i=1}^{N} \delta_i^{\Phi^{k-1}} t_i \varpi^{\mathrm{T}} h(x_i) \\ s.t. \; & t_i \varpi^{\mathrm{T}} h(x) \ge 1 - \frac{1}{1+\tau} \xi_i, \; \xi_i \ge 0, \; i = 1 \ldots N, \end{aligned} \tag{28}$$

where ϖ is the output weight vector in the ELM kernel space, and

$$\delta_i^{\Phi^{k-1}} = \begin{cases} cs_i^\Phi \tau, & t_i f_{\beta^{k-1}}(x_i) = t_i h(x_i) \varpi^{k-1} < \varsigma + 1 \\ 0, & \text{otherwise} \end{cases}. \tag{29}$$

The dual problem of (28) is follow as:

$$\begin{aligned} min_\alpha \; & \frac{1}{2} (\alpha^{\mathrm{T}} - \delta^{\mathrm{T}}) \widetilde{Q} (\alpha - \delta) - e^{\mathrm{T}} \alpha \\ s.t. \; & 0 \le \alpha \le (1+\tau) c S^\Phi, \end{aligned} \tag{30}$$

where $\widetilde{Q} = T \Omega_{ELM} T$ and $\Omega_{ELM} = H H^{\mathrm{T}} \in \Re^{N \times N}$ whose element $\Omega_{ELM ij} = \mathcal{K}_{ELM}(x_i, x_j)$.

Similar to linear TPin-IFELM, Eq. (30) is equivalent to

$$\begin{aligned} min_\lambda \; & \frac{1}{2} \lambda^\Phi \widetilde{Q} \lambda^\Phi - e^{\mathrm{T}} \alpha \\ s.t. \; & \mathfrak{L}^\Phi \le \lambda^\Phi \le \mathfrak{U}^\Phi, \end{aligned} \tag{31}$$

Algorithm 3 The complete process of nonlinear TPin-IFELM

Input:
$\{(x_i, y_i) \mid 1 \leq i \leq N\}$: the training set
c, ς, τ and K: the parameters
x: an unknown sample
Output:
t : the label of x
procedure :

(1) Obtain the score value s_i of each sample in the ELM kernel space according to Eq. (26).
(2) Obtain the optimal solution $\hat{\varpi}$ according to Algorithm 1.
(3) Determine the label t of the unknown sample x according to (32).

end procedure

where $\mathfrak{L}^{\Phi} = -\delta^{\Phi} \in \Re^N$ and $\mathfrak{U}^{\Phi} = (1+\tau)\, cS^{\Phi} - \delta^{\Phi} \in \Re^N$.

For the unknown sample x, the decision function of nonlinear TPin-IFELM is

$$f(x) = sign\left(\lambda^{\Phi^{\mathrm{T}}} T \begin{bmatrix} \mathcal{K}_{ELM}(x_1, x) \\ \vdots \\ \mathcal{K}_{ELM}(x_N, x) \end{bmatrix}\right). \tag{32}$$

The complete process of nonlinear TPin-IFELM is shown in Algorithm 3.

3.4 The Discussion

In this section, we discuss the relationship between TPin-IFELM and FELM. Similar to ELM, TPin-IFELM, and FELM randomly assign the input weights and the biases of the hidden layer. Then, the hidden layer output matrix is obtained by the activation function.

In order to suppress the negative effects of noises, FELM only uses the membership degree for each sample, while TPin-IFELM employs the membership and non-membership degrees based on the local information of samples. To further reduce the interference of noises, TPin-IFELM uses the truncated pinball loss function to not only maintain sparsity and robustness but also to enhance the re-sampling stability.

4 Properties of the TPin-IFELM

In this section, we analyze the theoretical properties of TPin-IFELM, including noise insensitivity, sparsity, weight scatter minimization, and misclassification error minimization.

4.1 Noise Insensitivity and Sparsity

In this subsection, we discuss the noise insensitivity and sparsity of TPin-IFELM. The subgradient function of (12) is

$$\partial P_{\tau,\varsigma}(1 - t_i f(x_i)) = \begin{cases} 0, & 1 - t_i f(x_i) < -\varsigma \\ [-\tau, 0], & 1 - t_i f(x_i) = -\varsigma \\ -\tau, & -\varsigma < 1 - t_i f(x_i) < 0 \\ [-\tau, 1], & 1 - t_i f(x_i) = 0 \\ 1, & 1 - t_i f(x_i) > 0 \end{cases}. \tag{33}$$

Equation (16) can be rewritten as:

$$\mathbf{0} \in \frac{\beta}{c} - \sum_{i=1}^{N} s_i t_i h\left(x_i\right) \partial P_{\tau,\varsigma}\left(1 - t_i h\left(x_i\right)^{\mathrm{T}} \beta\right), \tag{34}$$

where $\mathbf{0} \in \Re^N$ is a column vector whose elements are all zero.

For given β, the index set can be partitioned into five sets,

$$\begin{aligned}
\mathcal{S}_0^\beta &= \left\{i : 1 - t_i h\left(x_i\right)^{\mathrm{T}} \beta < -\varsigma\right\}, \\
\mathcal{S}_1^\beta &= \left\{i : 1 - t_i h\left(x_i\right)^{\mathrm{T}} \beta = -\varsigma\right\}, \\
\mathcal{S}_2^\beta &= \left\{i : -\varsigma < 1 - t_i h\left(x_i\right)^{\mathrm{T}} \beta < 0\right\}, \\
\mathcal{S}_3^\beta &= \left\{i : 1 - t_i h\left(x_i\right)^{\mathrm{T}} \beta = 0\right\}, \\
\mathcal{S}_4^\beta &= \left\{i : 1 - t_i h\left(x_i\right)^{\mathrm{T}} \beta > 0\right\}.
\end{aligned} \tag{35}$$

Since $\partial P_{\tau,\varsigma}\left(1 - t_i f\left(x_i\right)\right) = 0$ when the samples are located in $\mathcal{S}_0^\beta$, the samples in $\mathcal{S}_0^\beta$ have no contribution to the calculation of β. Therefore, $\mathcal{S}_0^\beta$ is closely related to the sparsity of (14), in the other word, the parameter ς can control the number of samples in $\mathcal{S}_0^\beta$. When the value of ς is smaller, the more samples in $\mathcal{S}_0^\beta$, and the better the sparsity of (14). In particular, when $\varsigma \to 0$, the truncated pinball loss function can be regarded as a hinge loss function, which is very sensitive to the noises. On the contrary, when the value of ς is larger, the number of samples in $\mathcal{S}_0^\beta$ is smaller, and (14) is robust to noises but gradually loses its sparsity. Particularly, when $\varsigma \to +\infty$, the truncated pinball loss function degenerates into pinball loss, and the sparsity is completely lost.

According to the five sets $\mathcal{S}_0^\beta$, $\mathcal{S}_1^\beta$, $\mathcal{S}_2^\beta$, $\mathcal{S}_3^\beta$ and $\mathcal{S}_4^\beta$ of (35), the optimality condition can be written as the existence of $\psi_i \in [-\tau, 0]$ and $\zeta_i \in [-\tau, 1]$ such that

$$\begin{aligned}
&\frac{\beta}{c} - \sum_{i \epsilon \mathcal{S}_1^\beta} \psi_i s_i t_i h\left(x_i\right) + \tau \sum_{i \epsilon \mathcal{S}_2^\beta} s_i t_i h\left(x_i\right) \\
&\quad - \sum_{i \epsilon \mathcal{S}_3^\beta} \zeta_i s_i t_i h\left(x_i\right) - \sum_{i \epsilon \mathcal{S}_4^\beta} s_i t_i h\left(x_i\right) = 0.
\end{aligned} \tag{36}$$

The number of samples in $\mathcal{S}_1^\beta$ and $\mathcal{S}_3^\beta$ is much smaller than that in $\mathcal{S}_0^\beta$, $\mathcal{S}_2^\beta$ and $\mathcal{S}_4^\beta$, and the samples in $\mathcal{S}_1^\beta$ and $\mathcal{S}_3^\beta$ make little contribution to Eq. (36). Therefore, the main problem here is about the set $\mathcal{S}_0^\beta$, $\mathcal{S}_2^\beta$ and $\mathcal{S}_4^\beta$. When the value of parameter ς is fixed to a suitable value, the parameter τ can be used to control the number of samples in $\mathcal{S}_0^\beta$, $\mathcal{S}_2^\beta$ and $\mathcal{S}_4^\beta$, and the sparsity of (14) is affected. When τ is large, such as $\tau = 1$, these three sets contain many samples, so (14) is robust to the feature noise. When τ is very small, such as $\tau = 0.1$, there are few samples in $\mathcal{S}_4^\beta$, and (14) is more sensitive. Especially, when $\tau = 0$, there are no samples or only a few samples in $\mathcal{S}_4^\beta$. Therefore, when constructing the model, the feature noise around the decision boundary will bring significant negative effects. Since the total number of samples is fixed when τ is smaller, the smaller the number of samples in $\mathcal{S}_4^\beta$, the larger the number of samples in $\mathcal{S}_0^\beta$, and the better the sparsity of (14).

In summary, the appropriate τ and ς are chosen to enable TPin-IFELM to better balance noise insensitivity and sparsity.

4.2 Weight Scatter and Misclassification Error Minimization

The mechanism of TPin-IFELM can also be explained by the weight scatter and misclassification error minimization. The positive hyperplane $f_+(x) : \beta^{\mathrm{T}} h(x) = 1$ and the negative hyperplane $f_-(x) : \beta^{\mathrm{T}} h(x) = -1$ are constructed by the samples in $\mathcal{S}_3^{\beta}$. The distance between positive and negative hyperplanes is $\frac{2}{\|\beta\|}$. We can measure the weight scatter in terms of the sum of the distances of a given point from similar samples. In the random mapping feature space related to β, the weight scatter of x_{i_0} can be defined as

$$\sum_{x_{i_0} \in \mathcal{S}_3^{\beta} \cap x_i \in \mathcal{X}} s_i |\beta^{\mathrm{T}} \left(h\left(x_{i_0}\right) - h\left(x_i\right)\right)|. \tag{37}$$

If $x_{i_0} \in \mathcal{S}_3^{\beta} \cap \mathcal{X}^+$, i.e., $\beta^{\mathrm{T}} h\left(x_{i_0}\right) = 1$ and $t_{i_0} = 1$, then

$$\sum_{x_i \in \mathcal{X}^+} s_i |\beta^{\mathrm{T}} \left(h\left(x_{i_0}\right) - h\left(x_i\right)\right)| = \sum_{x_i \in \mathcal{X}^+} s_i |1 - t_i \left(\beta^{\mathrm{T}} h\left(x_i\right)\right)|; \tag{38}$$

If $x_{i_0} \in \mathcal{S}_3^{\beta} \cap \mathcal{X}^-$, i.e., $\beta^{\mathrm{T}} h\left(x_{i_0}\right) = -1$ and $t_{i_0} = -1$, then

$$\sum_{x_i \in \mathcal{X}^-} s_i |\beta^{\mathrm{T}} \left(h\left(x_{i_0}\right) - h\left(x_i\right)\right)| = \sum_{x_i \in \mathcal{X}^-} s_i |1 - t_i \left(\beta^{\mathrm{T}} h\left(x_i\right)\right)|. \tag{39}$$

Therefore,

$$min_{\beta} \frac{1}{2} \|\beta\|^2 + C_1 \sum_{i=1}^{N} s_i |1 - t_i \left(\beta^{\mathrm{T}} h\left(x_i\right)\right)| \tag{40}$$

can be interpreted as maximizing the distance between hyperplanes $f_+(x)$ and $f_-(x)$ and meanwhile minimizing weight scatter.

In (14), (40) is extended to $P_{\tau,\varsigma}$. The misclassification term

$$C_1 min\left(s_i \left(1 - t_i \left(\beta^{\mathrm{T}} h\left(x_i\right)\right)\right), 0\right) - C_2 \left(L_{hinge}\left(s_i \left(1 - t_i \left(\beta^{\mathrm{T}} h\left(x_i\right)\right) + \varsigma\right)\right) - s_i \varsigma\right)$$

is introduced into Eq. (40), i.e.,

$$\begin{aligned} min_{\beta} & \frac{1}{2} \|\beta\|^2 + C_1 \sum_{i=1}^{N} s_i |1 - t_i \left(\beta^{\mathrm{T}} h\left(x_i\right)\right)| \\ & + C_1 \sum_{i=1}^{N} min\left(s_i \left(1 - t_i \left(\beta^{\mathrm{T}} h\left(x_i\right)\right)\right), 0\right) \\ & - C_2 \sum_{i=1}^{N} \left(L_{hinge}\left(s_i \left(1 - t_i \left(\beta^{\mathrm{T}} h\left(x_i\right)\right) + \varsigma\right)\right) - s_i \varsigma\right). \end{aligned} \tag{41}$$

We obtain TPin-IFELM with $C_1 = c(1 + \tau)$ and $C_2 = c\tau$. Thus, TPin-IFELM can minimize both the weight scatter and misclassification errors, simultaneously.

5 Experiments

In this section, we verify the effectiveness of TPin-IFELM through a series of experiments on the artificial dataset and benchmark datasets.[1]

5.1 Experimental Configuration

In order to evaluate the effectiveness of TPin-IFELM, we compared it with other eight advanced comparison algorithms. TPin-IFELM with SFA, which replaces the score function SV in TPin-IFELM with the score function SFA in PIFTSVM, contains four parameters c, L, τ, and ς, and TPin-IFELM contains five parameters c, L, τ, ς and K. To ensure the objectivity of the experiments, for the datasets with less than 2000 samples, the penalty parameter c of TPin-IFELM and TPin-IFELM with SFA, the penalty parameter C of OELM, RELM, and FELM, and the penalty parameters C_1 and C_2 of TELM, SPTELM, and PIFTSVM are searched from the set $\left\{2^i | i = -10, -8, \ldots, 8, 10\right\}$, and the number of hidden layer nodes L for these algorithms are searched from {50, 100, 200, 500}. For the datasets with greater than or equal to 2000 samples, the penalty parameters c, C, C_1 and C_2 are searched from $\left\{2^i | i = -10, -6, \ldots, 6, 10\right\}$, and the number of hidden layer nodes L is searched from {50, 100, 200}. τ and ς are searched from {0.25, 0.5, 0.75}, ε is searched from {0, 0.2, 0.5}, and for TPin-IFELM, the number of nearest neighbors K is searched from $\{1, 3, \ldots, 20\}$.

We implement all algorithms by using MATLAB (R2018a). The experimental environment is a workstation with the 11th Gen Intel Core i5-11,400 H (2.70GHz) processor and 16 G RAM. We use the quadprog to solve the quadratic programming problem and use three evaluation metrics to evaluate the classification performance, including accuracy (ACC), the area under ROC (AUC), and F_1-measure (F_1).

$$ACC = \frac{TP + TN}{TP + FN + TN + FP}, \tag{42}$$

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN}, \tag{43}$$

$$AUC = \frac{|\{(x_i, x_j) \mid f(x_j) \leq f(x_i), (x_i, x_j) \in P \times N\}|}{N^+ \times N^-}, \tag{44}$$

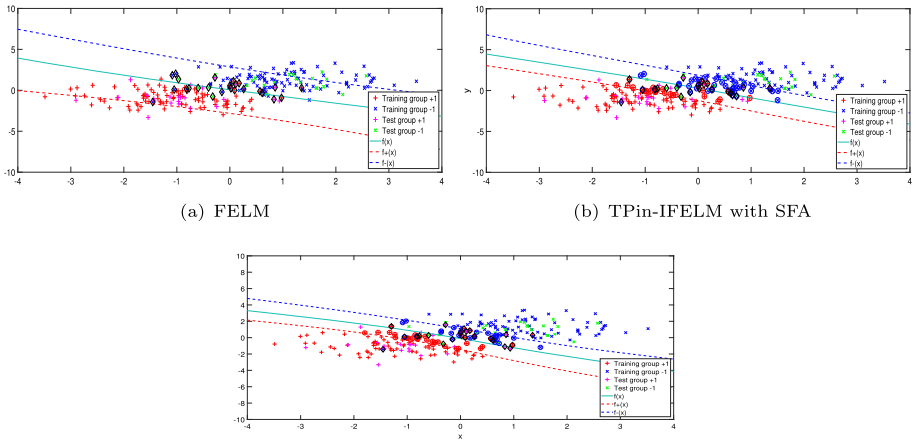
where FN denotes the number of false negatives, FP denotes the number of false positives, TN denotes the number of true negatives and TP denotes the number of true positives.

5.2 Experiments on the Artificial Dataset

To verify the robustness and sparsity of TPin-IFELM, we conduct comparative experiments on an artificial dataset with two-dimensional features. The training set and test sets consist of 200 samples and 50 samples, respectively. The positive and negative samples of the artificial dataset are generated by the Gaussian distributions $\mathcal{X}^+ \sim \mathcal{N}(\mathcal{V}_1, \Sigma_1)$ and $\mathcal{X}^- \sim \mathcal{N}(\mathcal{V}_2, \Sigma_2)$, respectively, where $\mathcal{V}_1 = \left[\begin{array}{cc} 1 & 1 \end{array}\right]^{\mathrm{T}}$, $\mathcal{V}_2 = \left[\begin{array}{cc} -1 & -1 \end{array}\right]^{\mathrm{T}}$ and $\Sigma_1 = \Sigma_2 = \left[\begin{array}{cc} 1 & \\ & 1 \end{array}\right]$.

As shown in Fig. 4, the red "+" and blue "×" denote the positive training samples and the negative training samples, respectively. The pink "+" and green "×" denote the positive test

[1] The datasets are available at https://archive.ics.uci.edu/ml/datasets.php and https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/.



(a) FELM (b) TPin-IFELM with SFA

(c) TPin-IFELM

Fig. 4 Separating boundaries of FELM, TPin-IFELM with SFA and TPin-IFELM

Table 1 The experimental results on artificial dataset

Evaluation metrics	FELM	TPin-IFELM with SFA	TPin-IFELM
ACC	0.8400	0.8400	**0.8600**
AUC	0.7056	0.7056	**0.7392**
F_1	0.8400	**0.8571**	0.8400

The best results for each dataset are shown in bold

samples and the negative test samples, respectively. The support vectors are circled by "○", and the noises identified by the algorithm are framed by black "◇". We can see that compared with FELM, TPin-IFELM with SFA and TPin-IFELM use both the membership and non-membership degrees and the truncated pinball loss function, so they can more effectively reduce the negative effect of noises. The number of support vectors of TPin-IFELM with SFA and TPin-IFELM is 33% and 29% of the total number of samples, respectively. Thus, compared with FELM, TPin-IFELM with SFA and TPin-IFELM are sparse. Table 1 shows the experimental results of FELM, TPin-IFELM with SFA, and TPin-IFELM on the artificial dataset, and the best results of each evaluation indicator are shown in bold. As shown in Table 1, TPin-IFELM is superior to FELM and TPin-IFELM with SFA in terms of ACC and AUC and is second only to TPin-IFELM with SFA in terms of F_1.

5.3 Experiments on the Benchmark Datasets

To evaluate the effectiveness and robustness of TPin-IFELM, we conduct comparative experiments on 15 benchmark datasets. The detailed characteristics of the datasets are shown in Table 2, where #Samples, #Positive samples, #Negative samples, and #Features denote the number of samples, the number of positive samples, the number of negative samples and the number of features, respectively.

In order to verify the classification performance of TPin-IFELM and other eight comparison algorithms, we conduct extensive experiments on fifteen benchmark datasets. Appendix B

Table 2 The characteristics of experimental datasets

Dataset	#Samples	#Positive samples	#Negative samples	#Features
Diabetes	768	500	268	8
WDBC	569	357	212	30
Australian	690	307	383	14
German	1000	300	700	24
Ionosphere	351	225	126	34
Heart	270	120	150	13
Madelon	2600	1300	1300	500
A3a	3185	773	2412	123
Splice	3175	1648	1527	60
Liver-disorders	345	155	190	5
Breast-cancer	683	444	239	10
Sonar	208	111	97	60
Svmguide3	1284	337	947	21
Colon-cancer	62	22	40	2000
A1a	1605	395	1210	123

provides additional experimental results. Unlike these seven comparison algorithms, TPin-IFELM with SFA and TPin-IFELM employ the membership and non-membership degrees to effectively identify the role of the samples in the classification process. As shown in Tables 7, 8 and 9, in terms of the average rank, each evaluation metric of TPin-IFELM is superior to that of the other eight algorithms, and the ACC and AUC of TPin-IFELM with SFA are only lower than TPin-IFELM.

Noise is commonly present in the datasets and can reduce the classification performance of algorithms. In order to demonstrate the robustness of TPin-IFELM, we conducted noise experiments on 15 benchmark datasets using label noise. We randomly select 50% of the training samples and then add label noise to them. The experimental results are shown in Tables 10, 11 and 12. We can observe that all algorithms are negatively affected by the samples with label noise. However, TPin-IFELM is less disturbed by label noise than the other eight comparison algorithms. In addition, it is superior to the other eight comparison algorithms on most datasets. For the classification problems with label noise and feature noise, we add Gaussian noise [39] that follows normal distribution $N\left(0, \sigma^2\right)$ to the training set to form a training set with feature noise, where σ is 0.5, and then randomly select 50% of the training samples as the samples with label noise. Tables 3, 4 and 5 show the experimental results. The best results for each dataset are shown in bold. TPin-IFELM with SFA and TPin-IFELM are less disturbed by label noise and feature noise than the other seven algorithms and are superior to them on most datasets.

From the above noise experimental results, we can observe that ELM, OELM, RELM, TELM, and SPTELM do not consider the membership degree of the samples to reduce the negative impact of the noise, resulting in a significant decrease in their classification performance. Different from FELM, TPin-IFELM with SFA and TPin-IFELM employ the membership and non-membership degrees to effectively identify the role of the samples and the noise in the classification process. At the same time, they introduce the truncated pinball loss function to enhance the robustness of the model. Compared to TPin-IFELM with

Table 3 ACC of nine algorithms in the 50% label noise and 0.5 Gaussian feature noise environment

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Diabetes	0.5766 ± 0.0421 (9)	0.6798 ± 0.0365 (4)	0.6518 ± 0.0271 (7)	0.6684 ± 0.0251 (6)	0.6812 ± 0.0108 (3)
	(500)	$(50, 2^{-6})$	$(50, 2^{-10})$	$(200, 2^{-6})$	$(500, 2^2, 2^2)$
WDBC	0.6126 ± 0.0334 (9)	0.7139 ± 0.0335 (7)	0.7653 ± 0.0657 (6)	0.7821 ± 0.0206 (5)	0.7951 ± 0.0161 (3)
	(50)	$(100, 2^6)$	$(50, 2^{10})$	$(200, 2^{-8})$	$(500, 2^{-6}, 2^{-6})$
Australian	0.5872 ± 0.0078 (9)	0.7261 ± 0.0180 (4)	0.7107 ± 0.0284 (5)	0.6994 ± 0.0300 (7)	0.7368 ± 0.0208 (3)
	(100)	$(200, 2^{-6})$	$(100, 2^{10})$	$(100, 2^{-8})$	$(500, 2^{-4}, 2^{-4})$
German	0.5680 ± 0.0182 (9)	0.6678 ± 0.0213 (6)	0.6666 ± 0.0474 (7)	0.6874 ± 0.0142 (4)	0.7058 ± 0.0004 (3)
	(100)	$(50, 2^{-10})$	$(50, 2^{-10})$	$(50, 2^{-10})$	$(50, 2^{-4}, 2^{-4})$
Ionosphere	0.6262 ± 0.0351 (9)	0.7186 ± 0.0388 (6)	0.6831 ± 0.0345 (8)	0.7293 ± 0.0219 (5)	0.7555 ± 0.0068 (3)
	(100)	$(50, 2^2)$	$(100, 2^0)$	$(50, 2^4)$	$(50, 2^2, 2^2)$
Heart	0.5770 ± 0.0002 (9)	0.6844 ± 0.0001 (6)	0.6600 ± 0.0003 (7)	0.6985 ± 0.0003 (5)	0.7215 ± 0.0002 (3)
	(100)	$(100, 2^{-10})$	$(50, 2^6)$	$(500, 2^{-8})$	$(50, 2^2, 2^2)$
Madelon	0.5295 ± 0.0001 (7)	0.5431 ± 0.0001 (5)	0.5365 ± 0.0000 (6)	0.5500 ± 0.0001 (3)	0.5138 ± 0.0001 (9)
	(100)	$(100, 2^{-6})$	$(500, 2^6)$	$(50, 2^{-10})$	$(50, 2^{-10}, 2^{-10})$
A3a	0.6198 ± 0.0004 (9)	0.7614 ± 0.0005 (4)	0.7642 ± 0.0002 (3)	0.7673 ± 0.0001 (2)	0.7507 ± 0.0002 (5)
	(50)	$(100, 2^{-10})$	$(200, 2^{-6})$	$(50, 2^{-10})$	$(100, 2^{10}, 2^{10})$
Splice	0.5759 ± 0.0001 (8)	0.6129 ± 0.0004 (5)	0.6150 ± 0.0000 (4)	0.6013 ± 0.0002 (6)	0.5597 ± 0.0003 (9)
	(500)	$(100, 2^6)$	$(50, 2^{-6})$	$(50, 2^{10})$	$(50, 2^{-6}, 2^{-6})$

Table 3 continued

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Liver-disorders	0.5525 ± 0.0286 (9)	0.6238 ± 0.0249 (7)	0.6307 ± 0.0132 (6)	0.6238 ± 0.0101 (8)	0.6609 ± 0.0005 (2)
	(200)	($100, 2^{10}$)	($200, 2^{-4}$)	($500, 2^{6}$)	($100, 2^{0}, 2^{0}$)
Breast-cancer	0.6535 ± 0.0193 (9)	0.8974 ± 0.0209 (4)	0.8970 ± 0.0379 (5)	0.8981 ± 0.0469 (3)	0.8755 ± 0.0002 (6)
	(100)	($50, 2^{-8}$)	($100, 2^{-6}$)	($500, 2^{-10}$)	($100, 2^{4}, 2^{4}$)
Sonar	0.6074 ± 0.0268 (9)	0.6673 ± 0.0052 (7)	0.6549 ± 0.0512 (8)	0.6814 ± 0.0304 (4)	0.6954 ± 0.0231 (3)
	(50)	($50, 2^{-6}$)	($500, 2^{0}$)	($100, 2^{-2}$)	($50, 2^{4}, 2^{4}$)
Svmguide3	0.5813 ± 0.0179 (9)	0.7246 ± 0.0252 (4)	0.7188 ± 0.0377 (5)	0.6992 ± 0.0175 (6)	0.7556 ± 0.0034 (2)
	(50)	($50, 2^{-10}$)	($200, 2^{-6}$)	($50, 2^{-10}$)	($200, 2^{-2}, 2^{-2}$)
Colon-cancer	0.6521 ± 0.0349 (8)	0.7872 ± 0.0089 (7)	0.7957 ± 0.0187 (4)	0.7885 ± 0.0487 (6)	0.8218 ± 0.0004 (3)
	(100)	($100, 2^{4}$)	($100, 2^{-4}$)	($200, 2^{-8}$)	($200, 2^{8}, 2^{8}$)
A1a	0.5626 ± 0.0002 (9)	0.7271 ± 0.0006 (7)	0.7533 ± 0.0003 (3.5)	0.7533 ± 0.0001 (3.5)	0.7464 ± 0.0003 (6)
	(200)	($100, 2^{-10}$)	($500, 2^{-4}$)	($200, 2^{-10}$)	($50, 2^{-10}, 2^{-10}$)
Average rank	8.7333	5.5333	5.6333	4.9000	4.2000

Datasets	SPTELM (L, C_1, C_2, ε)	PIFTSVM (C_1, C_2, τ, ε)	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM (L, C, τ, K, ς)
Diabetes	0.6188 ± 0.0218 (8)	0.6732 ± 0.0384 (5)	0.6953 ± 0.0340 (2)	**0.7300 ± 0.0122 (1)**
	($200, 2^{4}, 2^{4}$, 0.4)	($2^{8}, 2^{8}$, 0.75, 0)	($100, 2^{6}$, 0.25, 0.5)	($50, 2^{6}$, 0.75, 1, 0.5)
WDBC	0.6795 ± 0.0170 (8)	0.8683 ± 0.1231 (2)	0.7824 ± 0.1509 (4)	**0.9080 ± 0.0260 (1)**
	($100, 2^{10}, 2^{10}$, 0.4)	($2^{-4}, 2^{-4}$, 0.25, 0)	($50, 2^{-2}$, 0.75, 0.5)	($500, 2^{-4}$, 0.25, 1, 0.5)

Table 3 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Australian	0.6322 ± 0.0216 (8)	0.7029 ± 0.1000 (6)	$\mathbf{0.8043 \pm 0.0400}$ **(1)**	0.8017 ± 0.0481 (2)
	$(50, 2^{-10}, 2^{-10}, 0)$	$(2^0, 2^0, 0.25, 0)$	$(200, 2^{-4}, 0.75, 0.5)$	$(50, 2^{-4}, 0.25, 1, 0.5)$
German	0.6074 ± 0.0189 (8)	0.7090 ± 0.0114 (2)	0.6830 ± 0.0686 (5)	$\mathbf{0.7164 \pm 0.0023}$ **(1)**
	$(100, 2^2, 2^2, 0.2)$	$(2^0, 2^0, 0.25, 0)$	$(100, 2^{-8}, 0.5, 0.5)$	$(100, 2^{-2}, 0.25, 9, 0.5)$
Ionosphere	0.7134 ± 0.0301 (7)	0.7552 ± 0.0760 (4)	0.7777 ± 0.0606 (2)	$\mathbf{0.8086 \pm 0.0209}$ **(1)**
	$(500, 2^{-2}, 2^{-2}, 0)$	$(2^0, 2^0, 0.25, 0)$	$(100, 2^2, 0.75, 0.5)$	$(200, 2^{-4}, 0.25, 1, 0.5)$
Heart	0.6593 ± 0.0002 (8)	0.7148 ± 0.0985 (4)	0.7519 ± 0.0594 (2)	$\mathbf{0.7763 \pm 0.0002}$ **(1)**
	$(50, 2^{-4}, 2^{-4}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(200, 2^{-4}, 0.25, 0.5)$	$(50, 2^{-8}, 0.75, 1, 0.5)$
Madelon	0.5485 ± 0.0002 (4)	0.5223 ± 0.0230 (8)	0.5635 ± 0.0171 (2)	$\mathbf{0.5842 \pm 0.0001}$ **(1)**
	$(50, 2^{-6}, 2^{-6}, 0.2)$	$(2^6, 2^6, 0.75, 0)$	$(50, 2^{-2}, 0.75, 0.5)$	$(200, 2^6, 0.25, 3, 0.5)$
A3a	0.6465 ± 0.0004 (8)	0.7463 ± 0.0264 (6)	0.7366 ± 0.0887 (7)	$\mathbf{0.7796 \pm 0.0002}$ **(1)**
	$(50, 2^2, 2^2, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(200, 2^6, 0.25, 0.5)$	$(50, 2^{-10}, 0.25, 3, 0.5)$
Splice	0.6000 ± 0.0003 (7)	0.6224 ± 0.0146 (3)	0.6261 ± 0.0247 (2)	$\mathbf{0.6557 \pm 0.0002}$ **(1)**
	$(100, 2^{-6}, 2^{-6}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{-2}, 0.75, 0.5)$	$(50, 2^{-10}, 0.75, 19, 0.5)$
Liver-disorders	0.6470 ± 0.0075 (4)	0.6319 ± 0.0130 (5)	0.6580 ± 0.0487 (3)	$\mathbf{0.6887 \pm 0.0081}$ **(1)**
	$(50, 2^{10}, 2^{10}, 0)$	$(2^0, 2^0, 0.25, 0)$	$(500, 2^8, 0.75, 0.5)$	$(100, 2^2, 0.25, 1, 0.5)$
Breast-cancer	0.7323 ± 0.0203 (8)	0.8610 ± 0.1233 (7)	0.9473 ± 0.0373 (2)	$\mathbf{0.9570 \pm 0.0137}$ **(1)**
	$(50, 2^2, 2^2, 0.4)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{-2}, 0.25, 0.5)$	$(50, 2^{-4}, 0.5, 1, 0.5)$

Table 3 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Sonar	0.6741 ± 0.0149 (5)	0.6733± 0.0238 (6)	0.7167 ± 0.0705 (2)	**0.7731 ± 0.0212 (1)**
	$(50, 2^{10}, 2^{10}, 0.4)$	$(2^4, 2^4, 0.25, 0)$	$(50, 2^0, 0.75, 0.5)$	$(100, 2^4, 0.75, 1, 0.5)$
Svmguide3	0.6355 ± 0.0002 (8)	0.7546 ± 0.0267 (3)	0.6876 ± 0.0665 (7)	**0.7601 ± 0.0075 (1)**
	$(200, 2^{10}, 2^{10}, 0.4)$	$(2^2, 2^2, 0.25, 0)$	$(50, 2^2, 0.75, 0.5)$	$(200, 2^{-6}, 0.25, 1, 0.5)$
Colon-cancer	0.8577 ± 0.0010 (2)	0.6269± 0.1182 (9)	0.7910 ± 0.0394 (5)	**0.9359 ± 0.0007 (1)**
	$(50, 2^2, 2^2, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{-10}, 0.25, 0.5)$	$(100, 2^0, 0.5, 1, 0.5)$
A1a	0.6617 ± 0.0005 (8)	0.7551 ± 0.0319 (2)	0.7477 ± 0.0756 (5)	**0.7601 ± 0.0075 (1)**
	$(100, 2^2, 2^2, 0.4)$	$(2^{10}, 2^{10}, 0.25, 0)$	$(200, 2^{10}, 0.5, 0.5)$	$(50, 2^{-10}, 0.25, 3, 0.5)$
Average rank	6.7333	4.8000	3.4000	**1.0667**

The best results for each dataset are shown in bold

Table 4 AUC of nine algorithms in the 50% label noise and 0.5 Gaussian feature noise environment

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Diabetes	$0.5582 \pm 0.0467(9)$	$0.6077 \pm 0.0005(5)$	$0.5967 \pm 0.0263(6)$	$0.6160 \pm 0.0355(4)$	$0.6446 \pm 0.0328(2)$
	(500)	$(50, 2^{-6})$	$(50, 2^{-10})$	$(100, 2^{10})$	$(100, 2^{-10}, 2^{-10})$
WDBC	$0.6167 \pm 0.0119(9)$	$0.7613 \pm 0.0757(6)$	$0.7673 \pm 0.0391(5)$	$0.7342 \pm 0.0579(7)$	$0.7707 \pm 0.0194(4)$
	(50)	$(50, 2^{0})$	$(50, 2^{10})$	$(200, 2^{-8})$	$(500, 2^{-6}, 2^{-6})$
Australian	$0.6084 \pm 0.0327(9)$	$0.7290 \pm 0.0078(5)$	$0.7276 \pm 0.0201(6)$	$0.7446 \pm 0.0935(4)$	$0.7808 \pm 0.0415(3)$
	(100)	$(200, 2^{-6})$	$(100, 2^{10})$	$(50, 2^{-6})$	$(200, 2^{8}, 2^{8})$
German	$0.5691 \pm 0.0141(9)$	$0.6041 \pm 0.0227(6)$	$0.5993 \pm 0.0272(7)$	$0.5974 \pm 0.0354(8)$	$0.6077 \pm 0.0325(4)$
	(500)	$(50, 2^{6})$	$(500, 2^{10})$	$(100, 2^{-10})$	$(50, 2^{2}, 2^{2})$
Ionosphere	$0.5733 \pm 0.0296(9)$	$0.6987 \pm 0.0445(7)$	$0.6964 \pm 0.0397(8)$	$0.7055 \pm 0.0460(6)$	$0.7197 \pm 0.0543(3)$
	(100)	$(500, 2^{2})$	$(500, 2^{2})$	$(50, 2^{-10})$	$(50, 2^{2}, 2^{2})$
Heart	$0.6025 \pm 0.0208(9)$	$0.7753 \pm 0.0560(2)$	$0.6754 \pm 0.0177(6)$	$0.6570 \pm 0.1366(7)$	$0.7050 \pm 0.0792(5)$
	(100)	$(100, 2^{-10})$	$(500, 2^{2})$	$(500, 2^{-8})$	$(50, 2^{2}, 2^{2})$
Madelon	$0.5340 \pm 0.0267(8)$	$0.5564 \pm 0.0121(3)$	$0.5418 \pm 0.0236(6)$	$0.5474 \pm 0.0177(4)$	$0.5343 \pm 0.0206(7)$
	(50)	$(50, 2^{2})$	$(500, 2^{6})$	$(200, 2^{-10})$	$(100, 2^{6}, 2^{6})$
A3a	$0.5978 \pm 0.0537(9)$	$0.6632 \pm 0.0321(4)$	$0.6251 \pm 0.0638(6)$	$0.6725 \pm 0.0537(2)$	$0.6042 \pm 0.0700(7)$
	(500)	$(100, 2^{-6})$	$(200, 2^{-4})$	$(50, 2^{10})$	$(50, 2^{-2}, 2^{-2})$
Splice	$0.5753 \pm 0.0250(8)$	$0.6216 \pm 0.0403(5)$	$0.6155 \pm 0.0179(6)$	$0.6300 \pm 0.0448(2)$	$0.5423 \pm 0.0335(9)$
	(500)	$(100, 2^{6})$	$(200, 2^{-10})$	$(50, 2^{-10})$	$(100, 2^{-6}, 2^{-6})$

Table 4 continued

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Liver-disorders	0.6737 ± 0.0096(9)	0.7236 ± 0.0257(5)	0.7088 ± 0.0179(7)	0.7150 ± 0.0545(6)	0.7445 ± 0.0541(2)
	(500)	($200, 2^{-6}$)	($200, 2^{2}$)	($100, 2^{-10}$)	($50, 2^{4}, 2^{4}$)
Breast-cancer	0.6931 ± 0.0495(9)	0.8518 ± 0.0469(7)	0.8667 ± 0.0073(5)	0.8820 ± 0.0718(4)	0.9000 ± 0.0323(3)
	(100)	($200, 2^{-8}$)	($100, 2^{-6}$)	($500, 2^{-10}$)	($100, 2^{4}, 2^{4}$)
Sonar	0.3531 ± 0.0235(9)	0.4643 ± 0.0229(3)	0.4140 ± 0.0275(6)	0.4133 ± 0.0724(7)	0.4638 ± 0.0500(4)
	(50)	($100, 2^{-6}$)	($500, 2^{0}$)	($100, 2^{-2}$)	($50, 2^{4}, 2^{4}$)
Svmguide3	0.8312 ± 0.0151(9)	**1.0000 ± 0.0000(3.5)**	**1.0000 ± 0.0000(3.5)**	**1.0000 ± 0.0000(3.5)**	0.9980 ± 0.0044(7)
	(50)	($200, 2^{6}$)	($200, 2^{0}$)	($200, 2^{-2}$)	($100, 2^{-2}, 2^{-2}$)
Colon-cancer	0.6328 ± 0.0130(8)	0.7975 ± 0.0213(5)	0.8064 ± 0.0179(4)	0.7966 ± 0.0769(6)	0.8327 ± 0.1047(3)
	(100)	($200, 2^{4}$)	($100, 2^{-4}$)	($50, 2^{6}$)	($200, 2^{8}, 2^{8}$)
A1a	0.5945 ± 0.0573(8)	0.6548 ± 0.0346(5)	0.6435 ± 0.0575(6)	0.5879 ± 0.0494(9)	0.6629 ± 0.0222(4)
	(200)	($100, 2^{10}$)	($200, 2^{-4}$)	($50, 2^{10}$)	($50, 2^{6}, 2^{6}$)
Averagerank	8.7333	4.7667	5.8333	5.3000	4.4667

Datasets	SPTELM (L, C_1, C_2, ε)	PIFTSVM (C_1, C_2, τ, ε)	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM (L, C, τ, K, ς)
Diabetes	0.5799 ± 0.0554(8)	0.5920 ± 0.0842(7)	0.6248 ± 0.0613(3)	**0.6471 ± 0.0621(1)**
	($200, 2^{4}, 2^{4}$, 0.4)	($2^{8}, 2^{8}$, 0.75, 0)	($50, 2^{6}$, 0.25, 0.5)	($50, 2^{6}$, 0.75, 1, 0.5)
WDBC	0.6916 ± 0.0307(8)	0.8654 ± 0.1254(2)	0.7725 ± 0.1629(3)	**0.9324 ± 0.0567(1)**
	($100, 2^{10}, 2^{10}$, 0.4)	($2^{0}, 2^{0}$, 0.25, 0)	($50, 2^{-2}$, 0.75, 0.5)	($500, 2^{-4}$, 0.25, 1, 0.5)

Table 4 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Australian	$0.6567 \pm 0.0288(8)$	$0.7099 \pm 0.1382(7)$	$0.8279 \pm 0.0434(2)$	$\mathbf{0.8846 \pm 0.0527(1)}$
	$(50, 2^8, 2^8, 0)$	$(2^2, 2^2, 0.25, 0)$	$(200, 2^{-4}, 0.75, 0.5)$	$(50, 2^{-4}, 0.25, 1, 0.5)$
German	$0.6049 \pm 0.0380(5)$	$0.6141 \pm 0.0867(2)$	$0.6124 \pm 0.0447(3)$	$\mathbf{0.6354 \pm 0.0684(1)}$
	$(50, 2^{-10}, 2^{-10}, 0.2)$	$(2^4, 2^4, 0.25, 0)$	$(50, 2^8, 0.5, 0.5)$	$(50, 2^{10}, 0.25, 17, 0.5)$
Ionosphere	$0.7126 \pm 0.0864(4)$	$0.7067 \pm 0.0803(5)$	$0.7328 \pm 0.0639(2)$	$\mathbf{0.7426 \pm 0.0702(1)}$
	$(50, 2^{-2}, 2^{-2}, 0.4)$	$(2^4, 2^4, 0.25, 0)$	$(100, 2^2, 0.75, 0.5)$	$(200, 2^{-4}, 0.25, 1, 0.5)$
Heart	$0.6464 \pm 0.0429(8)$	$0.7378 \pm 0.0768(4)$	$0.7607 \pm 0.0825(3)$	$\mathbf{0.7969 \pm 0.0903(1)}$
	$(200, 2^{10}, 2^{10}, 0.4)$	$(2^2, 2^2, 0.75, 0)$	$(100, 2^{-4}, 0.25, 0.5)$	$(50, 2^{-8}, 0.75, 1, 0.5)$
Madelon	$0.5443 \pm 0.0260(5)$	$0.5281 \pm 0.0299(9)$	$0.5678 \pm 0.0076(2)$	$\mathbf{0.5948 \pm 0.0125(1)}$
	$(100, 2^{10}, 2^{10}, 0)$	$(2^6, 2^6, 0.25, 0)$	$(50, 2^{10}, 0.75, 0.5)$	$(50, 2^{-10}, 0.25, 3, 0.5)$
A3a	$0.6316 \pm 0.0534(5)$	$0.5990 \pm 0.0493(8)$	$0.6718 \pm 0.0378(3)$	$\mathbf{0.7222 \pm 0.0258(1)}$
	$(200, 2^{-2}, 2^{-2}, 0.4)$	$(2^{-2}, 2^{-2}, 0.75, 0)$	$(100, 2^{-6}, 0.75, 0.5)$	$(50, 2^{-10}, 0.25, 1, 0.5)$
Splice	$0.6119 \pm 0.0236(7)$	$0.6265 \pm 0.0133(4)$	$0.6290 \pm 0.0302(3)$	$\mathbf{0.6455 \pm 0.0184(1)}$
	$(100, 2^{10}, 2^{10}, 0)$	$(2^2, 2^2, 0.25, 0)$	$(50, 2^{-10}, 0.25, 0)$	$(50, 2^{-10}, 0.75, 19, 0.5)$
Liver-disorders	$0.7311 \pm 0.0864(4)$	$0.7052 \pm 0.0643(8)$	$0.7391 \pm 0.0617(3)$	$\mathbf{0.7669 \pm 0.0145(1)}$
	$(100, 2^4, 2^4, 0)$	$(2^0, 2^0, 0.25, 0)$	$(500, 2^8, 0.75, 0)$	$(100, 2^{-2}, 0.25, 1, 0.5)$
Breast-cancer	$0.7326 \pm 0.0617(8)$	$0.8579 \pm 0.1412(6)$	$\mathbf{0.9636 \pm 0.0276(1)}$	$0.9512 \pm 0.0498(2)$
	$(50, 2^2, 2^2, 0.4)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{-2}, 0.25, 0)$	$(100, 2^{-4}, 0.5, 1, 0.5)$

Table 4 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Sonar	$0.4558 \pm 0.0683(5)$	$0.4118 \pm 0.0734(8)$	$0.5071 \pm 0.1078(2)$	$\mathbf{0.5570 \pm 0.0990(1)}$
	$(50, 2^6, 2^6, 0.4)$	$(2^2, 2^2, 0.75, 0)$	$(50, 2^0, 0.75, 0)$	$(100, 2^4, 0.75, 1, 0.5)$
Svmguide3	$0.8706 \pm 0.0227(8)$	$\mathbf{1.0000 \pm 0.0000(3.5)}$	$\mathbf{1.0000 \pm 0.0000(3.5)}$	$\mathbf{1.0000 \pm 0.0000(3.5)}$
	$(200, 2^{10}, 2^0, 0.4)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{10}, 0.5, 0)$	$(50, 2^{-10}, 0.25, 1, 0.5)$
Colon-cancer	$0.8692 \pm 0.0705(2)$	$0.5544 \pm 0.0952(9)$	$0.7928 \pm 0.0722(7)$	$\mathbf{0.9822 \pm 0.0144(1)}$
	$(50, 2^2, 2^2, 0)$	$(2^{10}, 2^{10}, 0.5, 0)$	$(100, 2^{10}, 0.5, 0)$	$(100, 2^0, 0.5, 1, 0.5)$
A1a	$0.6277 \pm 0.0363(7)$	$0.6631 \pm 0.0844(3)$	$0.6748 \pm 0.0350(2)$	$\mathbf{0.7359 \pm 0.0482(1)}$
	$(50, 2^{-2}, 2^{-2}, 0.4)$	$(2^6, 2^6, 0.25, 0)$	$(50, 2^{-6}, 0.25, 0)$	$(50, 2^2, 0.75, 11, 0.5)$
Averagerank	6.1333	5.7000	2.8333	**1.2333**

The best results for each dataset are shown in bold

Table 5 F_1 of nine algorithms in the 50% label noise and 0.5 Gaussian feature noise environment

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Diabetes	0.6549 ± 0.0519 (9)	0.7630 ± 0.0073 (7)	0.7762 ± 0.0506 (6)	0.7913 ± 0.0236 (5)	0.7948 ± 0.0209 (3)
	(500)	$(50, 2^{-6})$	$(50, 2^{-10})$	$(200, 2^{-6})$	$(500, 2^2, 2^2)$
WDBC	0.5381 ± 0.0224 (9)	0.6953 ± 0.0919 (6)	0.6965 ± 0.0200 (5)	0.6566 ± 0.0551 (7)	0.7079 ± 0.0841 (4)
	(50)	$(100, 2^6)$	$(50, 2^{10})$	$(200, 2^{-8})$	$(500, 2^{-6}, 2^{-6})$
Australian	0.5810 ± 0.0023 (9)	0.6881 ± 0.0115 (6)	0.6886 ± 0.0426 (5)	0.7352 ± 0.0800 (4)	0.7555 ± 0.0466 (3)
	(100)	$(200, 2^{-6})$	$(100, 2^{10})$	$(100, 2^{-8})$	$(500, 2^{-4}, 2^{-4})$
German	0.4345 ± 0.0140 (9)	0.4695 ± 0.0281 (7)	0.4734 ± 0.0289 (5)	0.4773 ± 0.0597 (4)	0.4823 ± 0.0550 (3)
	(100)	$(50, 2^{-10})$	$(50, 2^{-10})$	$(50, 2^{-10})$	$(50, 2^{-4}, 2^{-4})$
Ionosphere	0.4897 ± 0.0377 (9)	0.6085 ± 0.0445 (8)	0.6186 ± 0.0253 (6)	0.6134 ± 0.0761 (7)	0.6260 ± 0.0719 (4)
	(100)	$(50, 2^2)$	$(100, 2^0)$	$(50, 2^4)$	$(50, 2^2, 2^2)$
Heart	0.6267 ± 0.0136 (9)	0.8091 ± 0.0361 (2)	0.7459 ± 0.0191 (5)	0.6813 ± 0.0743 (8)	0.7266 ± 0.0871 (6)
	(100)	$(100, 2^{-10})$	$(50, 2^6)$	$(500, 2^{-8})$	$(50, 2^2, 2^2)$
Madelon	0.4858 ± 0.0547 (9)	0.5633 ± 0.1135 (8)	0.5851 ± 0.0989 (7)	0.6469 ± 0.0555 (5)	0.6666 ± 0.0199 (4)
	(100)	$(100, 2^{-6})$	$(500, 2^6)$	$(50, 2^{-10})$	$(50, 2^{-10}, 2^{-10})$
A3a	0.4139 ± 0.0858 (9)	0.4914 ± 0.0360 (4)	0.4350 ± 0.0680 (7)	0.5045 ± 0.0733 (3)	0.4377 ± 0.0696 (6)
	(50)	$(100, 2^{-10})$	$(200, 2^{-6})$	$(50, 2^{-10})$	$(100, 2^{10}, 2^{10})$
Splice	0.5775 ± 0.0283 (9)	0.6527 ± 0.0627 (7)	0.6362 ± 0.0554 (8)	0.6833 ± 0.0292 (4)	0.6836 ± 0.0221 (3)
	(500)	$(100, 2^6)$	$(50, 2^{-6})$	$(50, 2^{10})$	$(50, 2^{-6}, 2^{-6})$

Table 5 continued

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Liver-disorders	0.5510 ± 0.0109 (9)	0.6019 ± 0.0238 (7)	0.6092 ± 0.0177 (6)	0.5954 ± 0.0604 (8)	0.6405 ± 0.0424 (2)
	(200)	($100, 2^{10}$)	($200, 2^{-4}$)	($500, 2^{6}$)	($100, 2^{0}, 2^{0}$)
Breast-cancer	0.6996 ± 0.0790 (9)	0.8671 ± 0.0550 (7)	0.8954 ± 0.0340 (6)	0.9096 ± 0.0603 (4)	0.9111 ± 0.0171 (3)
	(100)	($50, 2^{-8}$)	($100, 2^{-6}$)	($500, 2^{-10}$)	($100, 2^{4}, 2^{4}$)
Sonar	0.5836 ± 0.0169 (9)	0.6700 ± 0.0164 (6)	0.6624 ± 0.0206 (7)	0.6366 ± 0.0712 (8)	0.7190 ± 0.0210 (2)
	(50)	($50, 2^{-6}$)	($500, 2^{0}$)	($100, 2^{-2}$)	($50, 2^{4}, 2^{4}$)
Svmguide3	0.4130 ± 0.0024 (9)	0.4560 ± 0.0653 (5)	0.4496 ± 0.0398 (6)	0.4272 ± 0.0476 (8)	0.4572 ± 0.0430 (4)
	(50)	($50, 2^{-10}$)	($200, 2^{-6}$)	($50, 2^{-10}$)	($200, 2^{-2}, 2^{-2}$)
Colon-cancer	0.5826 ± 0.0193 (8)	0.7057 ± 0.0059 (7)	0.7455 ± 0.0152 (3)	0.7143 ± 0.0653 (6)	0.7821 ± 0.0427 (2)
	(100)	($100, 2^{4}$)	($100, 2^{-4}$)	($200, 2^{-8}$)	($200, 2^{8}, 2^{8}$)
A1a	0.4091 ± 0.0536 (8)	0.4749 ± 0.0769 (5)	0.4599 ± 0.0634 (6)	0.3962 ± 0.0569 (9)	0.4789 ± 0.0127 (4)
	(200)	($100, 2^{-10}$)	($500, 2^{4}$)	($200, 2^{-10}$)	($50, 2^{-10}, 2^{-10}$)
Average rank	8.8667	6.1333	5.8667	6.0000	3.5333

Datasets	SPTELM (L, C_1, C_2, ε)	PIFTSVM (C_1, C_2, τ, ε)	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM (L, C, τ, K, ς)
Diabetes	0.6898 ± 0.0400 (8)	0.7941 ± 0.0184 (4)	0.8009 ± 0.0347 (2)	**0.8039 ± 0.0619 (1)**
	($200, 2^{4}, 2^{4}$, 0.4)	($2^{8}, 2^{8}$, 0.75, 0)	($100, 2^{6}$, 0.25, 0)	($50, 2^{6}$, 0.75, 1, 0.5)
WDBC	0.6058 ± 0.0460 (8)	0.8356 ± 0.1185 (2)	0.7144 ± 0.2061 (3)	**0.8855 ± 0.0394 (1)**
	($100, 2^{10}, 2^{10}$, 0.4)	($2^{-4}, 2^{-4}$, 0.25, 0)	($50, 2^{-2}$, 0.75, 0)	($500, 2^{-4}$, 0.25, 1, 0.5)

Table 5 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Australian	0.6266 ± 0.0508 (8)	0.6877 ± 0.1301 (7)	0.7923 ± 0.0313 (2)	**0.8506 ± 0.0494 (1)**
	$(50, 2^{-10}, 2^{-10}, 0)$	$(2^0, 2^0, 0.25, 0)$	$(200, 2^{-4}, 0.75, 0)$	$(50, 2^{-4}, 0.25, 1, 0.5)$
German	0.4721 ± 0.0325 (6)	0.4969 ± 0.0441 (2)	0.4633 ± 0.0417 (8)	**0.5034 ± 0.0350 (1)**
	$(100, 2^2, 2^2, 0.2)$	$(2^0, 2^0, 0.25, 0)$	$(100, 2^{-8}, 0.5, 0)$	$(100, 2^{-2}, 0.25, 9, 0.5)$
Ionosphere	0.6343 ± 0.1049 (3)	0.6234 ± 0.0788 (5)	0.6492 ± 0.0356 (2)	**0.6643 ± 0.0647 (1)**
	$(500, 2^{-2}, 2^{-2}, 0)$	$(2^0, 2^0, 0.25, 0)$	$(100, 2^2, 0.75, 0)$	$(200, 2^{-4}, 0.25, 1, 0.5)$
Heart	0.6889 ± 0.0615 (7)	0.7761 ± 0.0707 (4)	0.7969 ± 0.0308 (3)	**0.8224 ± 0.1063 (1)**
	$(50, 2^{-4}, 2^{-4}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(200, 2^{-4}, 0.25, 0)$	$(50, 2^{-8}, 0.75, 1, 0.5)$
Madelon	0.6676 ± 0.0124 (2)	0.6666 ± 0.0090 (3)	0.6143 ± 0.0954 (6)	**0.6723 ± 0.0270 (1)**
	$(50, 2^{-6}, 2^{-6}, 0.2)$	$(2^6, 2^6, 0.75, 0)$	$(50, 2^{-2}, 0.75, 0)$	$(200, 2^6, 0.25, 3, 0.5)$
A3a	0.4553 ± 0.0609 (5)	0.4318 ± 0.0155 (8)	0.5053 ± 0.0468 (2)	**0.5483 ± 0.0208 (1)**
	$(50, 2^2, 2^2, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(200, 2^6, 0.25, 0)$	$(50, 2^{-10}, 0.25, 3, 0.5)$
Splice	0.6589 ± 0.0190 (6)	0.6866 ± 0.0104 (2)	0.6655 ± 0.0524 (5)	**0.6953 ± 0.0224 (1)**
	$(100, 2^{-6}, 2^{-6}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{-2}, 0.75, 0)$	$(50, 2^{-10}, 0.75, 19, 0.5)$
Liver-disorders	0.6289 ± 0.0326 (4)	0.6198 ± 0.0768 (5)	0.6333 ± 0.1017 (3)	**0.6712 ± 0.0334 (1)**
	$(50, 2^{10}, 2^{10}, 0)$	$(2^0, 2^0, 0.25, 0)$	$(500, 2^8, 0.75, 0)$	$(500, 2^2, 0.25, 1, 0.5)$
Breast-cancer	0.7585 ± 0.0578 (8)	0.9049 ± 0.0771 (5)	**0.9604 ± 0.0283 (1)**	0.9541 ± 0.0351 (2)
	$(50, 2^2, 2^2, 0.4)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{-2}, 0.25, 0)$	$(50, 2^{-4}, 0.5, 1, 0.5)$

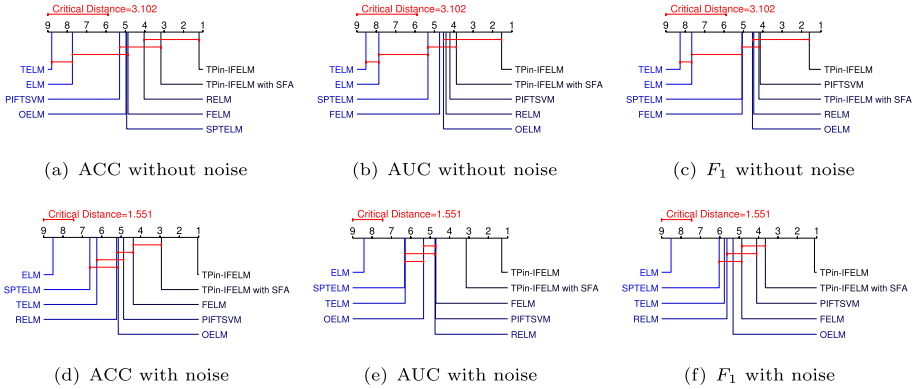
Table 5 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Sonar	0.6742 ± 0.0431 (5)	0.6762 ± 0.0623 (4)	0.7141 ± 0.0428 (3)	**0.7278 ± 0.0943 (1)**
	$(50, 2^{10}, 2^{10}, 0.4)$	$(2^4, 2^4, 0.25, 0)$	$(50, 2^0, 0.75, 0)$	$(100, 2^4, 0.75, 1, 0.5)$
Svmguide3	0.4601 ± 0.0287 (3)	0.4417 ± 0.0411 (7)	**0.4836 ± 0.0180 (1)**	0.4602 ± 0.0768 (2)
	$(200, 2^{10}, 2^{10}, 0.4)$	$(2^2, 2^2, 0.25, 0)$	$(50, 2^2, 0.75, 0)$	$(200, 2^{-6}, 0.25, 1, 0.5)$
Colon-cancer	0.7319 ± 0.1009 (4)	0.4986 ± 0.1749 (9)	0.7188 ± 0.0867 (5)	**0.9092 ± 0.0888 (1)**
	$(50, 2^2, 2^2, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{-10}, 0.25, 0)$	$(100, 2^0, 0.5, 1, 0.5)$
A1a	0.4442 ± 0.0513 (7)	0.4860 ± 0.0922 (3)	0.4972 ± 0.0285 (2)	**0.5686 ± 0.0550 (1)**
	$(100, 2^2, 2^2, 0.4)$	$(2^{10}, 2^{10}, 0.25, 0)$	$(200, 2^{10}, 0.25, 0)$	$(50, 2^2, 0.75, 11, 0.5)$
Average rank	5.6000	4.6667	3.2000	**1.1333**

The best results for each dataset are shown in bold

Table 6 Summary of Friedman statistics with and without noise and critical values with and without noise for all evaluation metrics

	Evaluation Metric	F_F	Critical Value ($\alpha = 0.05$)
Noise-free	ACC	29.6590	
	AUC	20.1525	2.0221
	F_1	15.2117	
Presence of noise	ACC	94.6143	
	AUC	69.7149	1.9580
	F_1	68.5880	



(a) ACC without noise (b) AUC without noise (c) F_1 without noise

(d) ACC with noise (e) AUC with noise (f) F_1 with noise

Fig. 5 Comparison results of TPinIFELM and eight comparison algorithms using the Nemenyi test on datasets without noise and datasets with noise

SFA, TPin-IFELM uses the local information of the samples to construct the more appropriate membership and non-membership degrees of the samples. Therefore, TPin-IFELM can better solve the classification problems with noise than the other eight comparison algorithms.

5.4 Statistical Analysis

From Tables 7, 8, 9, 10, 11 and 12 and Tables 3, 4 and 5, we can observe that not any algorithm outperforms all other algorithms on all datasets. In this subsection, we use the Friedman test [42] to analyze these algorithms statistically. Given $\mathfrak{K}$ comparison algorithms and $\mathcal{N}$ datasets, let r_i^j denote the rank of the j-th algorithm on the i-th dataset. $R_j = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} r_i^j$ denotes the average rank of the j-th algorithm. The Friedman statistics $F_F = \frac{(\mathcal{N}-1)\chi_F^2}{\mathcal{N}(\mathfrak{K}-1)-\chi_F^2} \sim F\left(\mathfrak{K}-1, (\mathfrak{K}-1)(\mathcal{N}-1)\right)$, where $\chi_F^2 = \frac{12\mathcal{N}}{\mathfrak{K}(\mathfrak{K}+1)}\left[\sum_{j=1}^{\mathfrak{K}} R_j^2 - \frac{\mathfrak{K}(\mathfrak{K}+1)^2}{4}\right]$. Table 6 shows the Friedman test results on the datasets without noise and datasets with noise. We observe that the Friedman statistics are much larger than the critical value, so the null hypothesis that all algorithms have the same classification performance is rejected, i.e., there is a significant difference in classification performance among the algorithms.

The difference between TPin-IFELM and the other eight algorithms is compared by using the Nemenyi test [42]. The average rank difference between pairs of algorithms is compared by the critical difference (CD), where $\text{CD} = q_\alpha \sqrt{\frac{\mathfrak{K}(\mathfrak{K}+1)}{6\mathcal{N}}}$. For the Nemenyi

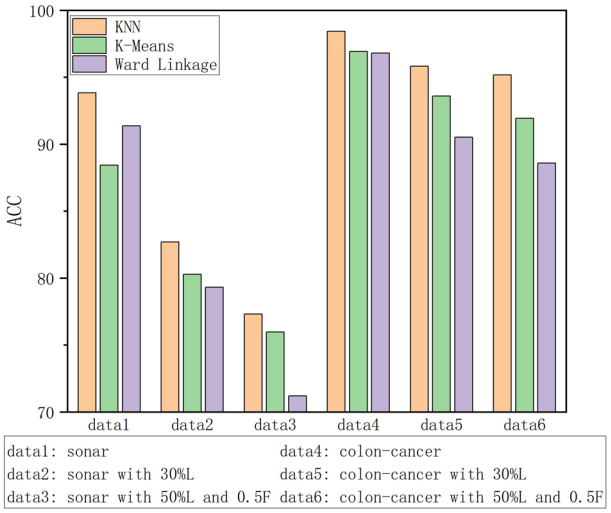


Fig. 6 Comparative results of three methods for obtaining sample structure information on datasets without noise and datasets with noise

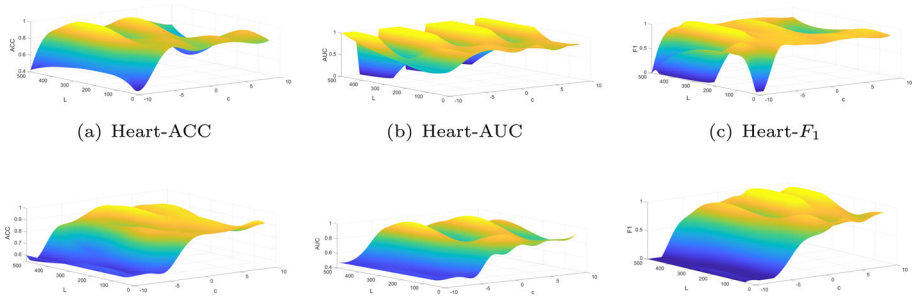


Fig. 7 Sensitivity analysis of parameters c and L on the datasets Heart and Ionosphere

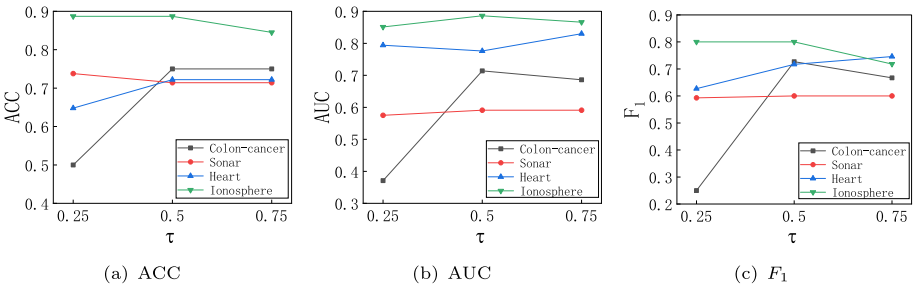
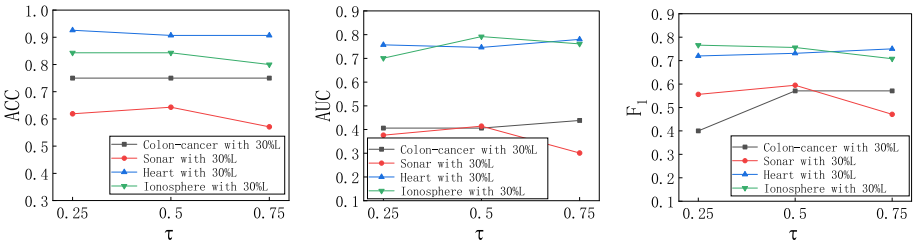
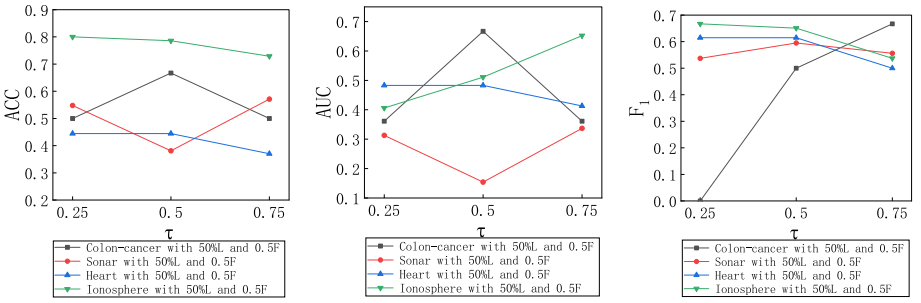


Fig. 8 The performance of TPin-IFELM on four datasets without noise changes with increasing the value of τ



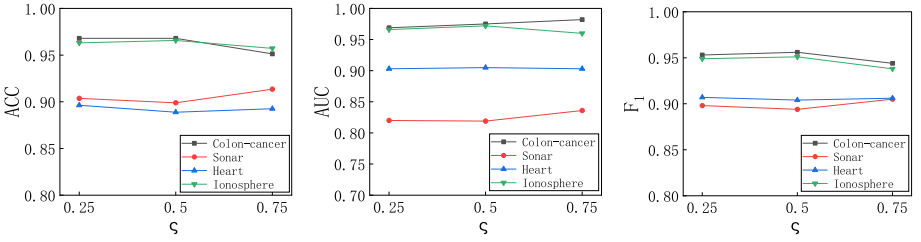
(a) ACC (b) AUC (c) F_1

Fig. 9 The performance of TPin-IFELM on four datasets with 30% label noise changes with increasing the value of τ



(a) ACC (b) AUC (c) F_1

Fig. 10 The performance of TPin-IFELM on four datasets with 50% label noise and feature noise of $\sigma = 0.5$ changes with increasing the value of τ



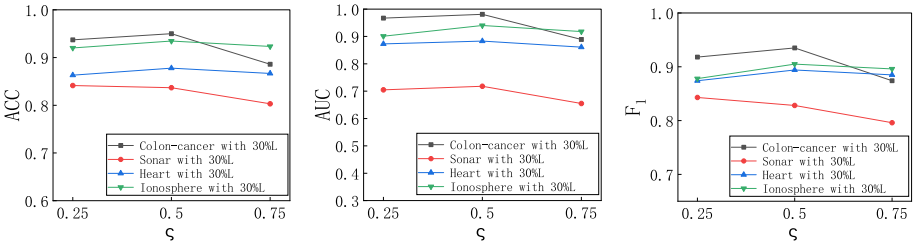
(a) ACC (b) AUC (c) F_1

Fig. 11 The performance of TPin-IFELM on four datasets without noise changes with increasing the value of ς

test, $q_\alpha = 2.948$ at the significance level $\alpha = 0.05$, thus, for experiments without noise, CD $= 3.102$ ($\Re = 9, \mathcal{N} = 15$); for experiments with noise, CD $= 1.5510$ ($\Re = 9, \mathcal{N} = 60$). The CD diagrams of all evaluation metrics with and without noise are shown in Fig. 5. We observe that TPin-IFELM is superior to the other eight algorithms on each evaluation metric.

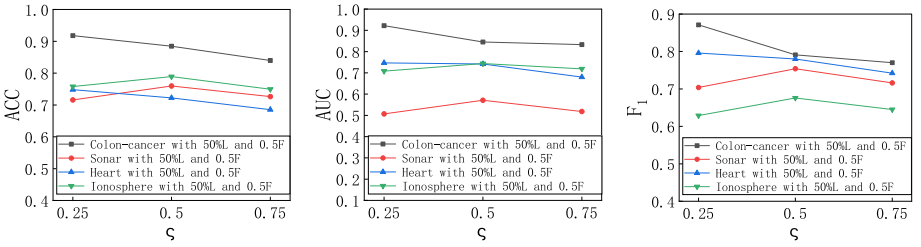
5.5 Sensitivity Analysis

To analyze the parameter sensitivity of TPin-IFELM and the performance of methods for obtaining sample structure information, we conduct experiments on the benchmark datasets.



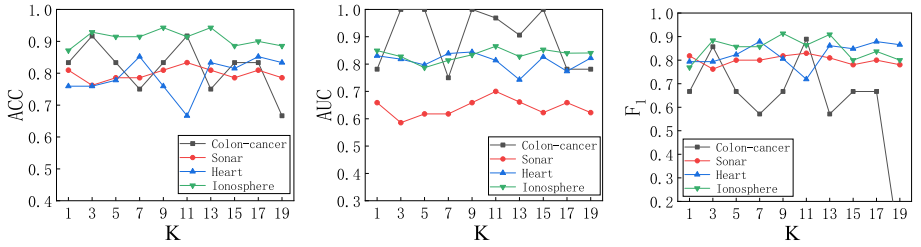
(a) ACC (b) AUC (c) F_1

Fig. 12 The performance of TPin-IFELM on four datasets with 30% label noise changes with increasing the value of ς



(a) ACC (b) AUC (c) F_1

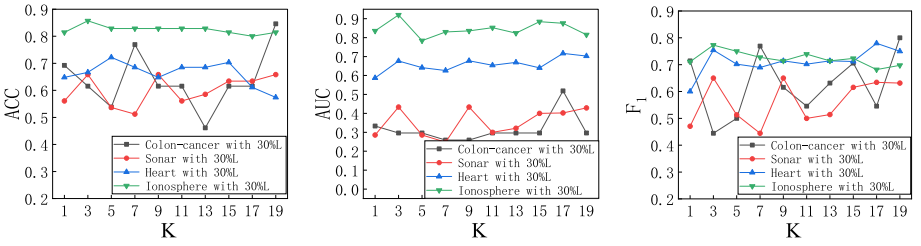
Fig. 13 The performance of TPin-IFELM on four datasets with 50% label noise and feature noise of $\sigma = 0.5$ changes with increasing the value of ς



(a) ACC (b) AUC (c) F_1

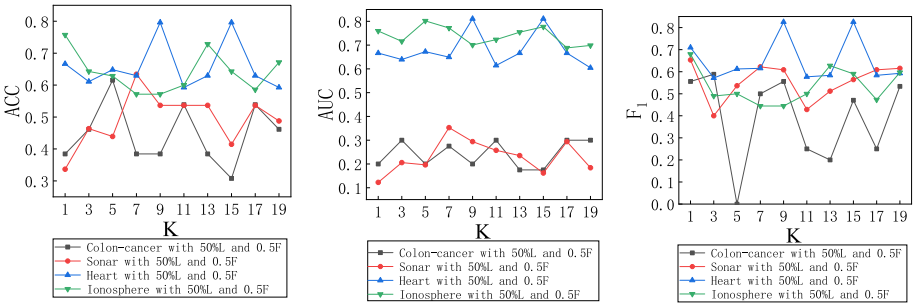
.

Fig. 14 The performance of TPin-IFELM on four datasets without noise changes with increasing the value of K



(a) ACC (b) AUC (c) F_1

Fig. 15 The performance of TPin-IFELM on four datasets with 30% label noise changes with increasing the value of K



(a) ACC (b) AUC (c) F_1

Fig. 16 The performance of TPin-IFELM on four datasets with 50% label noise and feature noise of $\sigma = 0.5$ changes with increasing the value of K

The main parameters of TPin-IFELM include the penalty parameter c, the number L of hidden layer nodes, the parameter τ, the parameter K, and the parameter ς. The methods for obtaining sample structure information include KNN, K-Means, and Ward Linkage.

5.5.1 Methods for Obtaining Sample Structure Information

In order to investigate the impact of different methods of obtaining sample structure information on our TPin-IFELM, we use KNN, K-Means and Ward Linkage to extract local information of samples, respectively, and conduct experiments on the datasets Sonar and Colon-cancer, respectively. The comparative results are shown in Fig. 6. As shown in Fig. 6, TPin-IFELM using KNN achieves optimal performance. Compared to K-Means and Ward Linkage, KNN can more effectively capture the correlation between the sample and all heterogeneous samples in its neighborhood, thus obtaining valuable local information of samples.

5.5.2 Parameters c and L

To analyze the sensitivity of TPin-IFELM to c and L, we perform parameter sensitivity analysis experiments on the datasets Heart and Ionosphere. The parameter c is searched form $\{2^i \mid i = -10, -8, \ldots, 8, 10\}$, the parameter L is searched form $\{50, 100, 200, 500\}$, and the other parameters are fixed. From Fig. 7, we can observe that the ACC, AUC, and F_1 of TPin-IFELM are higher when the value of c is larger and the L value is larger. In general, TPin-IFELM is sensitive to parameter c and is less affected by the change of L.

5.5.3 Parameters τ, ς and K

To analyze the effects of parameters τ, ς, and K on the classification performance of TPin-IFELM, we conducted experiments on the datasets Colon-cancer, Sonar, Heart and Ionosphere without noise and with noise, respectively. There are two types of noise: samples with 30% label noise and samples with 50% label noise and feature noise of $\sigma = 0.5$. As shown in Figs. 8, 9 and 10, for samples without noise, TPin-IFELM is minimally affected by the parameter τ, except for the data set Colon-cancer; however, for samples with noise, TPin-IFELM is strongly affected by the parameter τ. As shown in Figs. 11, 12 and 13, for samples without noise, TPin-IFELM is minimally affected by the parameter ς; however, for

samples with noise, TPin-IFELM is sensitive to the parameter ς. As shown in Figs. 14, 15 and 16, TPin-IFELM is sensitive to the parameter K.

6 Conclusion

Inspired by the intuitionistic fuzzy theory and truncated pinball loss, we propose a novel model to solve the classification problem in this paper. TPin-IFELM employs the KNN method to obtain the local information of samples, which can obtain the more suitable membership and non-membership degrees of samples. TPin-IFELM exploits the membership and non-membership degrees to effectively identify whether the boundary samples are noises or not and uses the truncated pinball loss function, which makes it more robust and sparse. A large number of experiments fully verify the effectiveness of TPin-IFELM. Compared with the state-of-the-art comparison algorithms, TPin-IFELM has superior classification performance. In future work, we will extend the proposed model to the multi-view classification problem.

Acknowledgements This work was supported in part by the Natural Science Foundation of Liaoning Province in China (2020-MS-281). We thank all anonymous reviewers for their helpful comments, which improved the quality of this paper.

Author Contributions Conceptualization: QG, QA; Methodology: QG, QA; Writing-original draft preparation: QG; Writing-review and editing: QG, QA; Funding acquisition: QA; Supervision: QA, WW.

Data availability The datasets generated during and/or analyzed during the current study are available in the UCI machine learning repository, https://archive.ics.uci.edu/ml/datasets.php, and the LIBSVM data repository, https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval All authors contributed to the conception and design of the study. All authors read and approved the final manuscript.

Appendix A Process of Obtaining the Dual Problem of (20)

In this section, we focus on solving the Problem (20). For clarity, the iteration superscript $k-1$ is removed. By introducing the Lagrangian multipliers α and θ, the Lagrangian function of the original problem (20) is

$$L(\beta, \xi, \alpha, \theta) = \frac{1}{2}\|\beta\|^2 + c\xi^{\mathrm{T}} S$$

$$+\delta^{\mathrm{T}} T H\beta - \alpha^{\mathrm{T}}\left(T H\beta - e + \frac{1}{1+\tau}\xi\right) - \theta^{\mathrm{T}}\xi, \tag{A1}$$

where $S=[s_1,\ldots,s_N]^{\mathrm{T}}, \delta=[\delta_1,\ldots,\delta_N]^{\mathrm{T}}, T=\begin{bmatrix} t_1 & & \\ & \ddots & \\ & & t_N \end{bmatrix}, H=[h(x_1),\ldots,h(x_N)]^{\mathrm{T}} \in \Re^{N\times L}$, $e=[1,\ldots,1]^{\mathrm{T}} \in \Re^{N}$, $\alpha=[\alpha_1,\ldots,\alpha_N]^{\mathrm{T}}$ and $\theta=[\theta_1,\ldots,\theta_N]^{\mathrm{T}}$ are the Lagrangian multiplier vectors.

According to the KKT conditions, we can obtain

$$\nabla_\beta L = \beta + H^{\mathrm{T}} T\delta - H^{\mathrm{T}} T\alpha = 0, \tag{A2}$$

$$\nabla_\xi L = cS - \frac{\alpha}{1+\tau} - \theta = 0, \tag{A3}$$

$$\alpha^{\mathrm{T}}\left(T H\beta - e + \frac{1}{1+\tau}\xi\right) = 0,\ \alpha \geq 0, \tag{A4}$$

$$t_i h(x_i)\beta \geq 1 - \frac{1}{1+\tau}\xi_i,\ \xi_i \geq 0, \tag{A5}$$

$$\theta^{\mathrm{T}}\xi = 0,\ \theta \geq 0. \tag{A6}$$

From (A2), we can obtain

$$\beta = H^{\mathrm{T}} T(\alpha - \delta). \tag{A7}$$

According to Eq. (A3) and Eq. (A6), we can derive

$$0 \leq \alpha \leq (1+\tau)cS. \tag{A8}$$

By substituting Eq. (A3) and Eq. (A7) into Eq. (A1), we can obtain the following dual problem.

$$\begin{aligned} &min_\alpha\ \frac{1}{2}(\alpha^{\mathrm{T}} - \delta^{\mathrm{T}})Q(\alpha - \delta) - e^{\mathrm{T}}\alpha \\ &s.t.\ 0 \leq \alpha \leq (1+\tau)cS, \end{aligned}$$

where $Q = THH^{\mathrm{T}}T$.

Appendix B Additional Experiments

We present the experimental results in the noise-free environment and 50% label noise environment. The best results for each dataset are shown in bold.

Table 7 ACC of nine algorithms in the noise-free environment

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Diabetes	0.7620 ± 0.0081 (8)	0.7883 ± 0.0029 (4)	0.7818 ± 0.0027 (7)	0.7820 ± 0.0062 (6)	0.6977 ± 0.0094 (9)
	(50)	$(50, 2^{-2})$	$(500, 2^{4})$	$(100, 2^{8})$	$(500, 2^{-6}, 2^{-6})$
WDBC	0.9631 ± 0.0037 (8)	0.9800 ± 0.0023 (5)	0.9838 ± 0.0015 (4)	0.9761 ± 0.0024 (7)	0.8046 ± 0.0202 (9)
	(200)	$(500, 2^{8})$	$(200, 2^{2})$	$(50, 2^{6})$	$(500, 2^{-8}, 2^{-8})$
Australian	0.8649 ± 0.0035 (8)	0.8832 ± 0.0033 (3)	0.8736 ± 0.0021 (7)	0.8739 ± 0.0018 (6)	0.7559 ± 0.0245 (9)
	(50)	$(200, 2^{-10})$	$(100, 2^{10})$	$(500, 2^{-6})$	$(500, 2^{6}, 2^{6})$
German	0.7628 ± 0.0055 (8)	0.7868 ± 0.0028 (3)	0.7798 ± 0.0022 (6)	0.7850 ± 0.0028 (4)	0.7154 ± 0.0053 (9)
	(50)	$(50, 2^{0})$	$(100, 2^{2})$	$(500, 2^{4})$	$(200, 2^{-4}, 2^{-4})$
Ionosphere	0.9003 ± 0.0149 (7)	0.9265 ± 0.0038 (6)	0.9470 ± 0.0026 (2)	0.9287 ± 0.0001 (5)	0.7532 ± 0.0202 (9)
	(50)	$(500, 2^{6})$	$(200, 2^{6})$	$(200, 2^{4})$	$(100, 2^{10}, 2^{10})$
Heart	0.7874 ± 0.0002 (8)	0.8652 ± 0.0056 (4)	0.8607 ± 0.0081 (5)	0.8593 ± 0.0041 (6)	0.7526 ± 0.0184 (9)
	(100)	$(50, 2^{-4})$	$(50, 2^{-4})$	$(100, 2^{6})$	$(200, 2^{4}, 2^{4})$
Madelon	0.5685 ± 0.0001 (9)	0.5979 ± 0.0000 (2)	0.5937 ± 0.0000 (4)	0.5951 ± 0.0001 (3)	0.5728 ± 0.0016 (8)
	(500)	$(200, 2^{-10})$	$(50, 2^{2})$	$(200, 2^{-6})$	$(50, 2^{-2}, 2^{-2})$
A3a	0.8380 ± 0.0002 (6)	0.8452 ± 0.0001 (2)	0.8352 ± 0.0001 (8)	0.8411 ± 0.0001 (4)	0.7582 ± 0.0001 (9)
	(200)	$(100, 2^{-2})$	$(200, 2^{-2})$	$(200, 2^{2})$	$(200, 2^{-10}, 2^{-10})$
Splice	0.8025 ± 0.0001 (8)	0.8501 ± 0.0035 (5)	0.8528 ± 0.0043 (3)	0.8523 ± 0.0041 (4)	0.6377 ± 0.0028 (9)
	(500)	$(500, 2^{-2})$	$(200, 2^{10})$	$(200, 2^{2})$	$(100, 2^{2}, 2^{2})$

Table 7 continued

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Liver-disorders	0.6110 ± 0.0095 (9)	0.6742 ± 0.0033 (5)	0.6649 ± 0.0102 (7)	0.6806 ± 0.0078 (4)	0.6464 ± 0.0004 (8)
	(50)	($50, 2^0$)	($50, 2^4$)	($50, 2^2$)	($200, 2^{-10}, 2^{-10}$)
Breast-cancer	0.9660 ± 0.0040 (8)	0.9778 ± 0.0012 (4)	0.9766 ± 0.0015 (5)	0.9754 ± 0.0024 (6)	0.8800 ± 0.0007 (9)
	(50)	($100, 2^4$)	($100, 2^6$)	($100, 2^8$)	($200, 2^{-6}, 2^{-6}$)
Sonar	0.8125 ± 0.0209 (8)	0.8740 ± 0.0094 (6)	0.8751 ± 0.0134 (5)	0.8894 ± 0.0089 (2)	0.7028 ± 0.0092 (9)
	(50)	($100, 2^4$)	($100, 2^6$)	($100, 2^8$)	($200, 2^{-6}, 2^{-6}$)
Svmguide3	0.8380 ± 0.0033 (6)	**0.8516 ± 0.0033 (1)**	0.8470 ± 0.0020 (3)	0.8464 ± 0.0018 (5)	0.7531 ± 0.0027 (9)
	(100)	($500, 2^{10}$)	($100, 2^8$)	($200, 2^{10}$)	($200, 2^{10}, 2^{10}$)
Colon-cancer	0.8547 ± 0.0141 (7)	0.9197 ± 0.0007 (6)	0.9308 ± 0.0100 (4)	0.9303 ± 0.0181 (5)	0.8244 ± 0.0006 (8)
	(500)	($100, 2^4$)	($200, 2^0$)	($500, 2^0$)	($50, 2^2, 2^2$)
A1a	0.8249 ± 0.0002 (8)	0.8430 ± 0.0001 (4.5)	0.8430 ± 0.0002 (4.5)	0.8411 ± 0.0002 (6)	0.7477 ± 0.0003 (9)
	(100)	($200, 2^{-6}$)	($100, 2^{-2}$)	($200, 2^6$)	($200, 2^{-10}, 2^{-10}$)
Average rank	7.7333	4.0333	4.9667	4.8667	8.8000

Datasets	SPTELM (L, C_1, C_2, ε)	PIFTSVM (C_1, C_2, τ, ε)	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM (L, C, τ, K, ς)
Diabetes	0.7904 ± 0.0023 (3)	0.7865 ± 0.0217 (5)	0.7904 ± 0.0495 (2)	**0.8016 ± 0.0027 (1)**
	($500, 2^0, 2^0$, 0)	($2^{-2}, 2^{-2}$, 0.25, 0)	($100, 2^8$, 0.25, 0.5)	($50, 2^0$, 0.25, 11, 0.5)
WDBC	0.9799 ± 0.0020 (6)	**0.9877 ± 0.0048 (1)**	0.9859 ± 0.0079 (2)	0.9845 ± 0.0015 (3)
	($50, 2^{-6}, 2^{-6}$, 0.2)	($2^{-4}, 2^{-4}$, 0.25, 0)	($50, 2^2$, 0.5, 0.5)	($50, 2^0$, 0.25, 7, 0.5)

Table 7 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Australian	0.8870 ± 0.0014 (2)	0.8783 ± 0.0188 (5)	0.8812 ± 0.0121 (4)	**0.8910 ± 0.0021 (1)**
	$(100, 2^{-2}, 2^{-2}, 0)$	$(2^{-4}, 2^{-4}, 0.75, 0)$	$(100, 2^{-8}, 0.75, 0.5)$	$(50, 2^{-6}, 0.5, 1, 0.5)$
German	0.7702 ± 0.0039 (7)	0.7830 ± 0.0309 (5)	0.7920 ± 0.0236 (2)	**0.8002 ± 0.0058 (1)**
	$(100, 2^{0}, 2^{0}, 0.4)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(200, 2^{2}, 0.25, 0.5)$	$(100, 2^{0}, 0.5, 1, 0.5)$
Ionosphere	0.9339 ± 0.0055 (4)	0.8946 ± 0.0409 (8)	0.9459 ± 0.0119 (3)	**0.9652 ± 0.0051 (1)**
	$(50, 2^{-2}, 2^{-2}, 0.4)$	$(2^{-6}, 2^{-6}, 0.75, 0)$	$(100, 2^{6}, 0.25, 0.5)$	$(50, 2^{6}, 0.25, 1, 0.5)$
Heart	0.8533 ± 0.0002 (7)	0.8667 ± 0.0401 (2.5)	0.8667 ± 0.0241 (2.5)	**0.8911 ± 0.0071 (1)**
	$(50, 2^{4}, 2^{4}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{0}, 0.5, 0.5)$	$(50, 2^{8}, 0.5, 1, 0.5)$
Madelon	0.5915 ± 0.0001 (6)	0.5754 ± 0.0247 (7)	0.5931 ± 0.0094 (5)	**0.6178 ± 0.0000 (1)**
	$(200, 2^{2}, 2^{2}, 0)$	$(2^{-2}, 2^{-2}, 0.75, 0)$	$(200, 2^{-2}, 0.25, 0.5)$	$(200, 2^{-6}, 0.5, 11, 0.5)$
A3a	0.8361 ± 0.0002 (7)	0.8433 ± 0.0083 (3)	0.8389 ± 0.0181 (5)	**0.8484 ± 0.0001 (1)**
	$(200, 2^{2}, 2^{2}, 0)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(200, 2^{2}, 0.25, 0.5)$	$(50, 2^{-2}, 0.25, 19, 0.5)$
Splice	0.8211 ± 0.0002 (7)	0.8476 ± 0.0184 (6)	0.8551 ± 0.0042 (2)	**0.8589 ± 0.0001 (1)**
	$(200, 2^{-10}, 2^{-10}, 0)$	$(2^{-2}, 2^{-2}, 0.75, 0)$	$(200, 2^{-2}, 0.25, 0.5)$	$(200, 2^{2}, 0.75, 19, 0.5)$
Liver-disorders	0.6672 ± 0.0090 (6)	0.6841 ± 0.0402 (3)	0.6899 ± 0.0194 (2)	**0.7177 ± 0.0033 (1)**
	$(100, 2^{0}, 2^{0}, 0)$	$(2^{-10}, 2^{-10}, 0.75, 0)$	$(50, 2^{-2}, 0.75, 0.5)$	$(100, 2^{-4}, 0.5, 1, 0.5)$
Breast-cancer	0.9798 ± 0.0028 (2)	0.9736 ± 0.0066 (7)	0.9795 ± 0.0096 (3)	**0.9804 ± 0.0013 (1)**
	$(100, 2^{-4}, 2^{-4}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(500, 2^{-2}, 0.25, 0.5)$	$(50, 2^{6}, 0.5, 1, 0.5)$

Table 7 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Sonar	0.8878 ± 0.0111 (3)	0.8273 ± 0.0558 (7)	0.8846 ± 0.0544 (4)	**0.9384 ± 0.0092 (1)**
	$(100, 2^{-4}, 2^{-4}, 0.2)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(500, 2^{2}, 0.5, 0.5)$	$(500, 2^{-2}, 0.75, 1, 0.5)$
Svmguide3	0.8466 ± 0.0003 (4)	0.8302 ± 0.0179 (8)	0.8341 ± 0.0255 (7)	0.8488 ± 0.0020 (2)
	$(50, 2^{-10}, 2^{-10}, 0.4)$	$(2^{-6}, 2^{-6}, 0.25, 0)$	$(200, 2^{10}, 0.5, 0.5)$	$(200, 2^{10}, 0.25, 3, 0.5)$
Colon-cancer	0.9205 ± 0.0009 (3)	0.3026 ± 0.1322 (9)	0.9346 ± 0.0694 (2)	**0.9842 ± 0.0160 (1)**
	$(200, 2^{-4}, 2^{-4}, 0.4)$	$(2^{10}, 2^{10}, 0.25, 0)$	$(100, 2^{0}, 0.75, 0.5)$	$(100, 2^{-2}, 0.25, 1, 0.5)$
A1a	0.8312 ± 0.0001 (7)	0.8461 ± 0.0384 (3)	0.8474 ± 0.0296 (2)	**0.8536 ± 0.0002 (1)**
	$(200, 2^{2}, 2^{2}, 0)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(100, 2^{6}, 0.25, 0.5)$	$(100, 2^{2}, 0.5, 1, 0.5)$
Average rank	4.9333	5.3000	3.1667	**1.2000**

The best results for each dataset are shown in bold

Table 8 AUC of nine algorithms in the noise-free environment

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Diabetes	0.7008 ± 0.0105 (8)	0.7201 ± 0.0099 (7)	0.7279 ± 0.0046 (6)	0.7281 ± 0.0536 (4)	0.6279 ± 0.0229 (9)
	(50)	(500, 2^4)	(500, 2^8)	(500, 2^8)	(500, 2^{-2}, 2^{-2})
WDBC	0.9644 ± 0.0092 (8)	0.9835 ± 0.0034 (5)	0.9862 ± 0.0027 (4)	0.9793 ± 0.0087 (7)	0.8113 ± 0.0678 (9)
	(200)	(500, 2^8)	(200, 2^2)	(100, 2^6)	(50, 2^2, 2^2)
Australian	0.8869 ± 0.0044 (8)	0.9006 ± 0.3337 (3)	0.8986 ± 0.0048 (5)	0.8990 ± 0.0323 (4)	0.7899 ± 0.0222 (9)
	(50)	(500, 2^0)	(500, 2^2)	(500, 2^{-8})	(500, 2^6, 2^6)
German	0.6968 ± 0.0082 (7)	0.7324 ± 0.0058 (5)	0.7277 ± 0.0034 (6)	0.7351 ± 0.0621 (3)	0.6141 ± 0.0306 (9)
	(200)	(50, 2^4)	(100, 2^2)	(100, 2^{-2})	(50, 2^0, 2^0)
Ionosphere	0.8686 ± 0.0095 (8)	0.9370 ± 0.0031 (2)	0.9352 ± 0.0010 (3)	0.9272 ± 0.0441 (5)	0.7238 ± 0.0493 (9)
	(50)	(50, 2^6)	(200, 2^6)	(200, 2^4)	(500, 2^{-6}, 2^{-6})
Heart	0.8048 ± 0.0136 (8)	0.8751 ± 0.0063 (4)	0.8756 ± 0.0051 (3)	0.8595 ± 0.0696 (6)	0.7389 ± 0.0672 (9)
	(100)	(50, 2^{-4})	(50, 2^{-4})	(100, 2^6)	(200, 2^4, 2^4)
Madelon	0.5674 ± 0.0179 (8)	0.5884 ± 0.0256 (3)	0.5757 ± 0.0037 (7)	0.5825 ± 0.0219 (4)	0.5453 ± 0.0452 (9)
	(500)	(200, 2^{-10})	(50, 2^2)	(200, 2^{-6})	(50, 2^{-2}, 2^{-2})
A3a	0.7375 ± 0.0179 (6)	0.7487 ± 0.0142 (5)	0.7601 ± 0.0075 (3)	0.7539 ± 0.0359 (4)	0.5980 ± 0.0604 (9)
	(200)	(100, 2^{-2})	(200, 2^6)	(200, 2^{10})	(50, 2^{-10}, 2^{-10})
Splice	0.8071 ± 0.0207 (8)	0.8285 ± 0.0169 (5)	0.8511 ± 0.0161 (3)	0.8188 ± 0.0176 (6)	0.5527 ± 0.0067 (9)
	(500)	(500, 2^{-2})	(200, 2^{10})	(200, 2^2)	(100, 2^2, 2^2)

Table 8 continued

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Liver-disorders	0.6893 ± 0.0050 (9)	0.7521 ± 0.0070 (6)	0.7485 ± 0.0044 (7)	0.7539 ± 0.0489 (5)	0.7215 ± 0.0723 (8)
	(50)	($50, 2^{-8}$)	($50, 2^{4}$)	($50, 2^{2}$)	($500, 2^{2}, 2^{2}$)
Breast-cancer	0.9778 ± 0.0023 (8)	0.9823 ± 0.0007 (4)	0.9821 ± 0.0006 (5)	0.9806 ± 0.0074 (6)	0.9045 ± 0.0520 (9)
	(50)	($100, 2^{2}$)	($50, 2^{8}$)	($100, 2^{8}$)	($200, 2^{-6}, 2^{-6}$)
Sonar	0.6653 ± 0.0419 (8)	0.7858 ± 0.0201 (2)	0.7738 ± 0.0050 (6)	0.7738 ± 0.0983 (5)	0.4056 ± 0.0350 (9)
	(500)	($500, 2^{10}$)	($500, 2^{4}$)	($500, 2^{10}$)	($500, 2^{-10}, 2^{-10}$)
Svmguide3	0.9777 ± 0.0006 (9)	**1.0000 ± 0.0000 (4)**	**1.0000 ± 0.0000 (4)**	**1.0000 ± 0.0000 (4)**	**1.0000 ± 0.0000 (4)**
	(100)	($500, 2^{-10}$)	($50, 2^{-10}$)	($500, 2^{10}$)	($50, 2^{6}, 2^{6}$)
Colon-cancer	0.8867 ± 0.0209 (7)	0.9358 ± 0.0022 (5)	0.9549 ± 0.0054 (3)	0.9405 ± 0.0305 (4)	0.8683 ± 0.1437 (8)
	(200)	($100, 2^{4}$)	($200, 2^{0}$)	($500, 2^{0}$)	($50, 2^{2}, 2^{2}$)
A1a	0.7365 ± 0.0442 (8)	0.7588 ± 0.0097 (6)	0.7714 ± 0.0337 (3)	0.7671 ± 0.0311 (4)	0.6767 ± 0.0496 (9)
	(200)	($500, 2^{4}$)	($50, 2^{8}$)	($200, 2^{2}$)	($50, 2^{2}, 2^{2}$)
Average rank	7.8667	4.4000	4.5333	4.7333	8.5333

Datasets	SPTELM (L, C_1, C_2, ε)	PIFTSVM ($C_1, C_2, \tau, \varepsilon$)	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM (L, C, τ, K, ς)
Diabetes	0.7321 ± 0.0256 (3)	**0.7393 ± 0.0241 (1)**	0.7280 ± 0.0657 (5)	0.7389 ± 0.0424 (2)
	($500, 2^{-4}, 2^{-4}$.0.25)	($2^{-10}, 2^{-10}$, 0.25, 0)	($100, 2^{8}$, 0.25, 0.5)	($50, 2^{2}$, 0.5, 9, 0.5)
WDBC	0.9813 ± 0.0201 (6)	0.9863 ± 0.0077 (3)	**0.9882 ± 0.0097 (1)**	0.9874 ± 0.0128 (2)
	($200, 2^{-10}, 2^{-10}$, 0.2)	($2^{-4}, 2^{-4}$, 0.25, 0)	($50, 2^{4}$, 0.25, 0.5)	($50, 2^{8}$, 0.25, 3, 0.5)

Table 8 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Australian	0.9045 ± 0.0342 (2)	0.8935 ± 0.0413 (7)	0.8978 ± 0.0101 (6)	**0.9091 ± 0.0349 (1)**
	$(200, 2^{-4}, 2^{-4}, 0)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(100, 2^{-8}, 0.75, 0.5)$	$(50, 2^{-6}, 0.5, 1, 0.5)$
German	0.6910 ± 0.0415 (8)	**0.7504 ± 0.0314 (1)**	0.7348 ± 0.0325 (4)	0.7477 ± 0.0416 (2)
	$(100, 2^{0}, 2^{0}, 0.4)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(200, 2^{4}, 0.25, 0.5)$	$(500, 2^{0}, 0.5, 1, 0.5)$
Ionosphere	0.9201 ± 0.0274 (6)	0.8687 ± 0.0411 (7)	0.9292 ± 0.0173 (4)	**0.9621 ± 0.0125 (1)**
	$(50, 2^{-10}, 2^{-10}, 0.4)$	$(2^{-6}, 2^{-6}, 0.75, 0)$	$(100, 2^{6}, 0.25, 0.5)$	$(200, 2^{10}, 0.5, 1, 0.5)$
Heart	0.8513 ± 0.0324 (7)	0.8772 ± 0.0223 (2)	0.8733 ± 0.0452 (5)	**0.8947 ± 0.0323 (1)**
	$(50, 2^{-4}, 2^{-4}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{0}, 0.5, 0.5)$	$(50, 2^{8}, 0.5, 1, 0.5)$
Madelon	0.5812 ± 0.0097 (5)	0.5780 ± 0.0305 (6)	0.5963 ± 0.0186 (2)	**0.6212 ± 0.0207 (1)**
	$(200, 2^{2}, 2^{2}, 0)$	$(2^{-2}, 2^{-2}, 0.75, 0)$	$(200, 2^{-2}, 0.25, 0.5)$	$(200, 2^{-6}, 0.5, 11, 0.5)$
A3a	0.7256 ± 0.0323 (7)	**0.8241 ± 0.0242 (1)**	0.7216 ± 0.0184 (8)	0.7793 ± 0.0173 (2)
	$(200, 2^{-2}, 2^{-2}, 0)$	$(2^{-2}, 2^{-2}, 0.75, 0)$	$(200, 2^{2}, 0.25, 0.5)$	$50, 2^{-6}, 0.25, 11, 0.5)$
Splice	0.8112 ± 0.0144 (7)	0.8459 ± 0.0190 (4)	0.8562 ± 0.0119 (2)	**0.8624 ± 0.0095 (1)**
	$(200, 2^{-10}, 2^{-10}, 0)$	$(2^{-2}, 2^{-2}, 0.75, 0)$	$(200, 2^{-2}, 0.25, 0.5)$	$(200, 2^{6}, 0.25, 19, 0.5)$
Liver-disorders	0.7632 ± 0.0466 (2)	0.7574 ± 0.0445 (3)	0.7543 ± 0.0258 (4)	**0.7917 ± 0.0465 (1)**
	$(500, 2^{-8}, 2^{-8}, 0.4)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{2}, 0.75, 0.5)$	$(100, 2^{-4}, 0.5, 1, 0.5)$
Breast-cancer	0.9836 ± 0.0112 (2)	0.9778 ± 0.0055 (7)	0.9830 ± 0.0048 (3)	**0.9855 ± 0.0114 (1)**
	$(100, 2^{-4}, 2^{-4}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{-8}, 0.25, 0.5)$	$(50, 2^{6}, 0.5, 1, 0.5)$

Table 8 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Sonar	0.7756 ± 0.0503 (4)	0.6888 ± 0.0847 (7)	0.7858 ± 0.1026 (3)	**0.8677 ± 0.0845 (1)**
	$(500, 2^{-6}, 2^{-6}, 0.2)$	$(2^{-4}, 2^{-4}, 0.25, 0)$	$(500, 2^{2}, 0.5, 0.5)$	$(500, 2^{-2}, 0.75, 1, 0.5)$
Svmguide3	0.9944 ± 0.0063 (8)	**1.0000 ± 0.0000 (4)**	**1.0000 ± 0.0000 (4)**	**1.0000 ± 0.0000 (4)**
	$(100, 2^{2}, 2^{2}, 0.4)$	$(2^{2}, 2^{2}, 0.25, 0)$	$(50, 2^{-10}, 0.25, 0.5)$	$(50, 2^{-10}, 0.25, 1, 0.5)$
Colon-cancer	0.9296 ± 0.0777 (6)	0.2674 ± 0.1098 (9)	0.9878 ± 0.0169 (2)	**0.9922 ± 0.0097 (1)**
	$(200, 2^{-4}, 2^{-4}, 0.4)$	$(2^{10}, 2^{10}, 0.25, 0)$	$(100, 2^{0}, 0.75, 0.5)$	$(100, 2^{-2}, 0.25, 1, 0.5)$
A1a	0.7366 ± 0.0220 (7)	**0.8308 ± 0.0366 (1)**	0.7627 ± 0.0425 (5)	0.7961 ± 0.0358 (2)
	$(200, 2^{-2}, 2^{-2}, 0)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(100, 2^{6}, 0.25, 0.5)$	$(100, 2^{6}, 0.25, 15, 0.5)$
Average rank	5.3333	4.2000	3.8667	**1.5333**

The best results for each dataset are shown in bold

Table 9 F_1 of nine algorithms in the noise-free environment

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Diabetes	0.8237 ± 0.0058 (8)	0.8493 ± 0.0235 (3)	0.8482 ± 0.0019 (5)	0.8425 ± 0.0241 (7)	0.8050 ± 0.0278 (9)
	(50)	$(500, 2^{-2})$	$(500, 2^{-4})$	$(100, 2^{8})$	$(500, 2^{-8}, 2^{-8})$
WDBC	0.9471 ± 0.0127 (8)	0.9756 ± 0.0045 (5)	0.9782 ± 0.0018 (4)	0.9710 ± 0.0060 (7)	0.7487 ± 0.0646 (9)
	(200)	$(500, 2^{8})$	$(200, 2^{2})$	$(100, 2^{6})$	$(50, 2^{2}, 2^{2})$
Australian	0.8569 ± 0.0059 (8)	0.8689 ± 0.0056 (4)	0.8690 ± 0.0019 (3)	0.8720 ± 0.0266 (2)	0.7601 ± 0.0212 (9)
	(50)	$(500, 2^{0})$	$(100, 2^{10})$	$(500, 2^{-8})$	$(200, 2^{4}, 2^{4})$
German	0.5370 ± 0.0101 (8)	0.6007 ± 0.0062 (5)	0.5878 ± 0.0058 (6)	0.6089 ± 0.0555 (3)	0.4886 ± 0.0408 (9)
	(200)	$(50, 2^{4})$	$(100, 2^{2})$	$(100, 2^{-2})$	$(200, 2^{10}, 2^{10})$
Ionosphere	0.8427 ± 0.0090 (7)	0.9159 ± 0.0052 (4)	0.9184 ± 0.0068 (3)	0.9057 ± 0.0259 (5)	0.6542 ± 0.0719 (9)
	(50)	$(500, 2^{6})$	$(200, 2^{6})$	$(200, 2^{4})$	$(500, 2^{-6}, 2^{-6})$
Heart	0.8257 ± 0.0060 (8)	0.8853 ± 0.0067 (2)	0.8807 ± 0.0106 (5)	0.8648 ± 0.0413 (7)	0.7401 ± 0.0566 (9)
	(100)	$(200, 2^{-4})$	$(500, 2^{-4})$	$(100, 2^{6})$	$(200, 2^{-10}, 2^{-10})$
Madelon	0.5673 ± 0.0383 (9)	0.6118 ± 0.0193 (8)	0.6506 ± 0.0173 (7)	0.6574 ± 0.0213 (5)	0.6667 ± 0.0157 (4)
	(50)	$(500, 2^{-10})$	$(100, 2^{2})$	$(500, 2^{-10})$	$(100, 2^{-2}, 2^{-2})$
A3a	0.6203 ± 0.0393 (6)	0.6372 ± 0.0338 (4)	0.6369 ± 0.0213 (5)	0.6456 ± 0.0490 (3)	0.4275 ± 0.0580 (9)
	(200)	$(200, 2^{-2})$	$(100, 2^{6})$	$(200, 2^{10})$	$(100, 2^{10}, 2^{10})$
Splice	0.8114 ± 0.0193 (8)	0.8332 ± 0.0231 (5)	0.8543 ± 0.0138 (3)	0.8233 ± 0.0144 (6)	0.6841 ± 0.0226 (9)
	(500)	$(500, 2^{-2})$	$(200, 2^{10})$	$(200, 2^{-2})$	$(50, 2^{-6}, 2^{-6})$

Table 9 continued

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Liver-disorders	0.5458 ± 0.0069 (9)	0.5968 ± 0.0222 (7)	0.5905 ± 0.0125 (8)	0.6040 ± 0.0369 (6)	0.6258 ± 0.0801 (4)
	(50)	($200, 2^0$)	($50, 2^4$)	($50, 2^2$)	($500, 2^2, 2^2$)
Breast-cancer	0.9767 ± 0.0023 (8)	0.9829 ± 0.0017 (3)	0.9802 ± 0.0015 (6)	0.9819 ± 0.0106 (5)	0.9133 ± 0.0461 (9)
	(50)	($100, 2^4$)	($100, 2^6$)	($100, 2^8$)	($200, 2^{-6}, 2^{-6}$)
Sonar	0.7851 ± 0.0222 (8)	0.8784 ± 0.0123 (3)	0.8709 ± 0.0029 (5)	0.8638 ± 0.0611 (6)	0.6623 ± 0.0693 (9)
	(500)	($500, 2^{10}$)	($500, 2^4$)	($500, 2^{10}$)	($100, 2^{-6}, 2^{-6}$)
Svmguide3	0.6373 ± 0.0046 (6)	0.6531 ± 0.0251 (3)	**0.6694 ± 0.0607 (1)**	0.6345 ± 0.0435 (7)	0.4796 ± 0.0574 (9)
	(100)	($500, 2^{10}$)	($500, 2^{10}$)	($200, 2^{10}$)	($10, 2^{10}, 2^{10}$)
Colon-cancer	0.7805 ± 0.0408 (7)	0.8699 ± 0.0548 (6)	0.8884 ± 0.0100 (4)	0.9013 ± 0.0296 (3)	0.7100 ± 0.1041 (8)
	(500)	($100, 2^4$)	($200, 2^0$)	($500, 2^0$)	($50, 2^2, 2^2$)
A1a	0.6000 ± 0.0705 (7)	0.6452 ± 0.0183 (5)	0.6531 ± 0.0599 (3)	0.6509 ± 0.0531 (4)	0.4977 ± 0.0577 (9)
	(200)	($500, 2^8$)	($50, 2^8$)	($200, 2^6$)	($50, 2^2, 2^2$)
Average rank	7.6667	4.4667	4.5333	5.0667	8.2667

Datasets	SPTELM (L, C_1, C_2, ε)	PIFTSVM ($C_1, C_2, \tau, \varepsilon$)	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM (L, C, τ, K, ς)
Diabetes	0.8485 ± 0.0248 (4)	0.8426 ± 0.0255 (6)	0.8524 ± 0.0385 (2)	**0.8574 ± 0.0224 (1)**
	($500, 2^2, 2^2, 0.2$)	($2^{-4}, 2^{-4}, 0.25, 0$)	($100, 2^8, 0.25, 0.5$)	($50, 2^6, 0.5, 3, 0.5$)
WDBC	0.9751 ± 0.0196 (6)	**0.9830 ± 0.0075 (1)**	0.9812 ± 0.0092 (2)	0.9810 ± 0.0115 (3)
	($50, 2^{-6}, 2^{-6}, 0.2$)	($2^{-2}, 2^{-2}, 0.25, 0$)	($50, 2^6, 0.25, 0.5$)	($50, 2^0, 0.25, 9, 0.5$)

Table 9 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Australian	0.8687 ± 0.0413 (5)	0.8676 ± 0.0372 (6)	0.8672 ± 0.0197 (7)	**0.8753 ± 0.0522 (1)**
	$(100, 2^{-2}, 2^{-2}, 0)$	$(2^{-4}, 2^{-4}, 0.25, 0)$	$(100, 2^{10}, 0.75, 0.5)$	$(50, 2^{-6}, 0.5, 3, 0.5)$
German	0.5392 ± 0.0474 (7)	**0.6255 ± 0.0594 (1)**	0.6057 ± 0.0374 (4)	0.6157 ± 0.0453 (2)
	$(100, 2^{0}, 2^{0}, 0.4)$	$(2^{-6}, 2^{-6}, 0.75, 0)$	$(200, 2^{4}, 0.25, 0.5)$	$(500, 2^{0}, 0.5, 3, 0.5)$
Ionosphere	0.8960 ± 0.0207 (6)	0.8347 ± 0.0452 (8)	0.9184 ± 0.0177 (2)	**0.9288 ± 0.0104 (1)**
	$(50, 2^{-2}, 2^{-2}, 0.4)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{6}, 0.25, 0.5)$	$(200, 2^{4}, 0.25, 3, 0.5)$
Heart	0.8658 ± 0.0524 (6)	0.8849 ± 0.0332 (3)	0.8834 ± 0.0262 (4)	**0.9028 ± 0.0178 (1)**
	$(50, 2^{2}, 2^{2}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{0}, 0.5, 0.5)$	$(50, 2^{8}, 0.5, 3, 0.5)$
Madelon	0.6674 ± 0.0140 (2)	0.6668 ± 0.0198 (3)	0.6571 ± 0.0183 (6)	**0.6737 ± 0.0217 (1)**
	$(200, 2^{6}, 2^{6}, 0.4)$	$(2^{2}, 2^{2}, 0.25, 0)$	$(50, 2^{-6}, 0.25, 0.5)$	$(200, 2^{10}, 0.75, 11, 0.5)$
A3a	0.5949 ± 0.0389 (8)	**0.6741 ± 0.0294 (1)**	0.6016 ± 0.0283 (7)	0.6616 ± 0.0233 (2)
	$(100, 2^{-2}, 2^{-2}, 0.4)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(200, 2^{2}, 0.25, 0.5)$	$(100, 2^{10}, 0.25, 13, 0.5)$
Splice	0.8182 ± 0.0160 (7)	0.8453 ± 0.0236 (4)	0.8563 ± 0.0059 (2)	**0.8607 ± 0.0123 (1)**
	$(200, 2^{-10}, 2^{-10}, 0)$	$(2^{-2}, 2^{-2}, 0.75, 0)$	$(200, 2^{-2}, 0.25, 0.5)$	$(200, 2^{2}, 0.75, 21, 0.5)$
Liver-disorders	0.6505 ± 0.0307 (2)	0.6409 ± 0.0415 (3)	0.6147 ± 0.0403 (5)	**0.6679 ± 0.0510 (1)**
	$(100, 2^{0}, 2^{0}, 0)$	$(2^{2}, 2^{2}, 0.75, 0)$	$(500, 2^{6}, 0.75, 0.5)$	$(500, 2^{8}, 0.25, 3, 0.5)$
Breast-cancer	0.9820 ± 0.0109 (4)	0.9798 ± 0.0042 (7)	0.9839 ± 0.0084 (2)	**0.9853 ± 0.0034 (1)**
	$(100, 2^{-4}, 2^{-4}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(500, 2^{-2}, 0.25, 0.5)$	$(50, 2^{6}, 0.5, 3, 0.5)$

Table 9 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Sonar	0.8800 ± 0.0293 (2)	0.8222 ± 0.0602 (7)	0.8784 ± 0.0557 (4)	**0.9186 ± 0.0558 (1)**
	$(50, 2^{-6}, 2^{-6}, 0.2)$	$(2^{0}, 2^{-10}, 0.25, 0)$	$(500, 2^{2}, 0.5, 0.5)$	$(500, 2^{-2}, 0.75, 3, 0.5)$
Svmguide3	0.6513 ± 0.0929 (4)	0.6597 ± 0.0284 (2)	0.5846 ± 0.0510 (8)	0.6393 ± 0.0312 (5)
	$(500, 2^{-10}, 2^{-10}, 0.4)$	$(2^{-6}, 2^{-6}, 0.25, 0)$	$(50, 2^{10}, 0.5, 0.5)$	$(200, 2^{10}, 0.25, 5, 0.5)$
Colon-cancer	0.8810 ± 0.0905 (5)	0.1371 ± 0.1300 (9)	0.9085 ± 0.0924 (2)	**0.9763 ± 0.0146 (1)**
	$(200, 2^{-4}, 2^{-4}, 0.4)$	$(2^{10}, 2^{10}, 0.25, 0)$	$(100, 2^{0}, 0.75, 0.5)$	$(100, 2^{-2}, 0.25, 3, 0.5)$
A1a	0.5954 ± 0.0431 (8)	**0.7030 ± 0.0613 (1)**	0.6380 ± 0.0367 (6)	0.6806 ± 0.0287 (2)
	$(200, 2^{-2}, 2^{-2}, 0)$	$(2^{-2}, 2^{-2}, 0.6, 0)$	$(100, 2^{6}, 0.27, 0.5)$	$(100, 2^{6}, 0.25, 15, 0.5)$
Average rank	5.0667	4.1333	4.2000	**1.6000**

The best results for each dataset are shown in bold

Table 10 ACC of nine algorithms in the 50% label noise environment

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Diabetes	0.5680 ± 0.0068 (9)	0.6368 ± 0.0362 (6)	0.6169 ± 0.0609 (7)	0.6726 ± 0.0077 (5)	0.6865 ± 0.0097 (4)
	(50)	$(100, 2^{8})$	$(100, 2^{-2})$	$(100, 2^{10})$	$(500, 2^{4}, 2^{4})$
WDBC	0.5905 ± 0.0263 (9)	0.7420 ± 0.0214 (7)	0.7911 ± 0.0561 (5)	0.7741 ± 0.0449 (6)	0.7972 ± 0.0324 (4)
	(50)	$(100, 2^{-4})$	$(500, 2^{-2})$	$(100, 2^{-6})$	$(200, 2^{2}, 2^{2})$
Australian	0.5858 ± 0.0260 (9)	0.6684 ± 0.0181 (7)	0.7235 ± 0.0440 (5)	0.6942 ± 0.0187 (6)	0.7490 ± 0.0089 (4)
	(500)	$(100, 2^{-4})$	$(200, 2^{0})$	$(50, 2^{-2})$	$(500, 2^{-6}, 2^{-6})$
German	0.5690 ± 0.0152 (9)	0.6316 ± 0.0186 (7)	0.6726 ± 0.0241 (6)	0.6962 ± 0.0124 (5)	0.7090 ± 0.0072 (4)
	(200)	(50, 10)	$(50, 2^{-10})$	$(50, 2^{-8})$	$(200, 2^{0}, 2^{0})$
Ionosphere	0.6256 ± 0.0234 (9)	0.6848 ± 0.0178 (7)	0.6813 ± 0.0237 (8)	0.6946 ± 0.0305 (6)	0.7612 ± 0.0089 (3)
	(200)	$(500, -6)$	$(200, 2^{10})$	$(100, 2^{-4})$	$(100, 2^{0}, 2^{0})$
Heart	0.5654 ± 0.0001 (9)	0.6274 ± 0.0002 (8)	0.6467 ± 0.0003 (7)	0.7000 ± 0.0002 (5)	0.7170 ± 0.0001 (3)
	(50)	$(500, 2^{-6})$	$(50, 2^{8})$	$(50, 2^{10})$	$(500, 2^{-10}, 2^{-10})$
Madelon	0.5218 ± 0.0000 (8)	0.5571 ± 0.0000 (5)	0.5561 ± 0.0001 (6)	0.5579 ± 0.0001 (4)	0.5173 ± 0.0001 (9)
	(200)	$(50, 2^{-6})$	$(500, 2^{0})$	$(500, 2^{-2})$	$(100, 2^{10}, 2^{10})$
A3a	0.5953 ± 0.0004 (9)	0.7721 ± 0.0002 (3)	0.7598 ± 0.0001 (4)	0.7473 ± 0.0003 (6)	0.7454 ± 0.0002 (7)
	(50)	$(50, 2^{-10})$	$(50, 2^{-10})$	$(200, 2^{-10})$	$(200, 2^{-2}, 2^{-2})$
Splice	0.5673 ± 0.0001 (9)	0.6120 ± 0.0003 (5)	0.6321 ± 0.0002 (3)	0.6167 ± 0.0002 (4)	0.5691 ± 0.0004 (8)
	(200)	$(50, 2^{-6})$	$(500, 2^{-6})$	$(50, 2^{2})$	$(50, 2^{2}, 2^{2})$

Table 10 continued

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Liver-disorders	0.5623 ± 0.0206 (9)	0.6093 ± 0.0097 (8)	0.6226 ± 0.0168 (6)	0.6203 ± 0.0143 (7)	0.6609 ± 0.0003 (2.5)
	(500)	$(500, 2^{-6})$	$(500, 2^{8})$	$(100, 2^{2})$	$(500, 2^{6}, 2^{6})$
Breast-cancer	0.6665 ± 0.0228 (9)	0.8415 ± 0.0363 (7)	0.8923 ± 0.0287 (4)	0.8764 ± 0.0364 (5)	0.8682 ± 0.0003 (6)
	(50)	$(50, 2^{2})$	$(50, 2^{-4})$	$(200, 2^{-8})$	$(100, 2^{-2}, 2^{-2})$
Sonar	0.6060 ± 0.0226 (9)	0.6346 ± 0.0137 (8)	0.6606 ± 0.0123 (5)	0.6491 ± 0.0271 (6)	0.6894 ± 0.0096 (3)
	(50)	$(100, 2^{-2})$	$(500, 2^{-2})$	$(50, 2^{2})$	$(50, 2^{2}, 2^{2})$
Svmguide3	0.6061 ± 0.0264 (9)	0.6972 ± 0.0185 (7)	0.7090 ± 0.0203 (6)	0.7316 ± 0.0160 (5)	0.7489 ± 0.0056 (3)
	(50)	$(50, 2^{-10})$	$(50, 2^{-10})$	$(50, 2^{-10})$	$(50, 2^{-10}, 2^{-10})$
Colon-cancer	0.6406 ± 0.0075 (8)	0.7692 ± 0.0377 (7)	0.8021 ± 0.0255 (3)	0.7846 ± 0.0080 (6)	0.7897 ± 0.0008 (4)
	(100)	$(50, 2^{2})$	$(50, 2^{-2})$	$(50, 2^{10})$	$(100, 2^{-10}, 2^{-10})$
A1a	0.6087 ± 0.0005 (9)	0.7458 ± 0.0004 (7)	0.7545 ± 0.0003 (4)	0.7564 ± 0.0001 (3)	0.7470 ± 0.0003 (6)
	(50)	$(50, 2^{-10})$	$(200, 2^{-6})$	$(50, 2^{-10})$	$(100, 2^{10}, 2^{10})$
Average rank	8.8667	6.6000	5.2667	5.2667	4.7000

Datasets	SPTELM (L, C_1, C_2, ε)	PIFTSVM ($C_1, C_2, \tau, \varepsilon$)	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM (L, C, τ, K, ς)
Diabetes	0.6151 ± 0.0066 (8)	0.6940 ± 0.0303 (2)	0.6887 ± 0.0490 (3)	**0.7141 ± 0.0226 (1)**
	$(50, 2^{-8}, 2^{-8}, 0.4)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{-2}, 0.25, 0.5)$	$(200, 2^{8}, 0.5, 5, 0.5)$
WDBC	0.6766 ± 0.0254 (8)	0.8050 ± 0.0798 (3)	0.8279 ± 0.0739 (2)	**0.8773 ± 0.0288 (1)**
	$(200, 2^{10}, 2^{10}, 0.4)$	$(2^{2}, 2^{2}, 0.25, 0)$	$(50, 2^{8}, 0.5, 0.5)$	$(50, 2^{0}, 0.25, 11, 0.5)$

Table 10 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Australian	0.6377 ± 0.0174 (8)	0.7507 ± 0.0707 (3)	0.7870 ± 0.0819 (2)	**0.8330 ± 0.0200 (1)**
	$(100, 2^{-2}, 2^{-2}, 0.2)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(50, 2^{0}, 0.25, 0.5)$	$(50, 2^{-2}, 0.5, 15, 0.5)$
German	0.6024 ± 0.0054 (8)	0.7170 ± 0.0372 (2)	**0.7230 ± 0.0271 (1)**	0.7154 ± 0.0068 (3)
	$(100, 2^{2}, 2^{2}, 0.2)$	$(2^{4}, 2^{4}, 0.25, 0)$	$(50, 2^{-10}, 0.25, 0.5)$	$(200, 2^{-6}, 0.5, 5, 0.5)$
Ionosphere	0.6968 ± 0.0192 (5)	0.7722 ± 0.0684 (2)	0.7551 ± 0.1030 (4)	**0.7859 ± 0.0157 (1)**
	$(100, 2^{2}, 2^{2}, 0.4)$	$(2^{2}, 2^{2}, 0.25, 0)$	$(50, 2^{4}, 0.75, 0.5)$	$(200, 2^{6}, 0.5, 15, 0.5)$
Heart	0.6667 ± 0.0002 (6)	0.7185 ± 0.0685 (2)	0.7000 ± 0.0757 (4)	**0.7511 ± 0.0001 (1)**
	$(200, 2^{-10}, 2^{-10}, 0.2)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(200, 2^{2}, 0.5, 0.5)$	$(100, 2^{4}, 0.5, 11, 0.5)$
Madelon	0.5615 ± 0.0003 (3)	0.5365 ± 0.0095 (7)	0.5658 ± 0.0048 (2)	**0.5898 ± 0.0000 (1)**
	$(50, 2^{-2}, 2^{-2}, 0.4)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(200, 2^{-2}, 0.5, 0.5)$	$(50, 2^{-2}, 0.5, 1, 0.5)$
A3a	0.6534 ± 0.0002 (8)	0.7520 ± 0.0240 (5)	0.7724 ± 0.0297 (2)	**0.7780 ± 0.0003 (1)**
	$(50, 2^{2}, 2^{2}, 0)$	$(2^{2}, 2^{2}, 0.25, 0)$	$(200, 2^{2}, 0.75, 0.5)$	$(50, 2^{-10}, 0.25, 13, 0.5)$
Splice	0.5912 ± 0.0002 (7)	0.5972 ± 0.0399 (6)	0.6457 ± 0.0215 (2)	**0.6724 ± 0.0001 (1)**
	$(50, 2^{-6}, 2^{-6}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{-6}, 0.5, 0.5)$	$(50, 2^{2}, 0.75, 9, 0.5)$
Liver-disorders	0.6296 ± 0.0075 (5)	0.6377 ± 0.0703 (4)	0.6609 ± 0.0318 (2.5)	**0.7032 ± 0.0129 (1)**
	$(50, 2^{6}, 2^{6}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(200, 2^{0}, 0.75, 0.5)$	$(50, 2^{6}, 0.5, 13, 0.5)$
Breast-cancer	0.7320 ± 0.0234 (8)	0.9296 ± 0.0590 (2)	0.9107 ± 0.0460 (3)	**0.9575 ± 0.0111 (1)**
	$(50, 2^{2}, 2^{2}, 0.2)$	$(2^{-4}, 2^{-4}, 0.25, 0)$	$(50, 2^{-10}, 0.25, 0.5)$	$(50, 2^{-2}, 0.25, 5, 0.5)$

Table 10 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Sonar	0.6713 ± 0.0139 (4)	0.6441 ± 0.0549 (7)	0.6973 ± 0.0876 (2)	**0.7471 ± 0.0269 (1)**
	$(200, 2^8, 2^8, 0)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(100, 2^4, 0.5, 0.5)$	$(50, 2^8, 0.75, 7, 0.5)$
Svmguide3	0.6404 ± 0.0133 (8)	0.7531 ± 0.0323 (2)	0.7414 ± 0.0404 (4)	**0.7575 ± 0.0126 (1)**
	$(200, 2^2, 2^2, 0.2)$	$(2^2, 2^2, 0.25, 0)$	$(50, 2^2, 0.75, 0.5)$	$(50, 2^2, 0.75, 3, 0.5)$
Colon-cancer	0.7885 ± 0.0005 (5)	0.6269 ± 0.2015 (9)	0.8538 ± 0.0391 (2)	**0.9047 ± 0.0160 (1)**
	$(100, 2^{-10}, 2^{-10}, 0.2)$	$(2^{10}, 2^{10}, 0.25, 0)$	$(50, 2^8, 0.25, 0.5)$	$(50, 2^2, 0.75, 15, 0.5)$
A1a	0.6542 ± 0.0005 (8)	0.7526 ± 0.0149 (5)	0.7813 ± 0.0184 (2)	**0.7819 ± 0.0003 (1)**
	$(50, 2^2, 2^2, 0.2)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(200, 2^6, 0.5, 0.5)$	$(50, 2^{-10}, 0.25, 13, 0.5)$
Average rank	6.6000	4.0667	2.5000	**1.1333**

The best results for each dataset are shown in bold

Table 11 AUC of nine algorithms in the 50% label noise environment

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Diabetes	0.5742 ± 0.0196 (9)	0.6092 ± 0.0278 (6)	0.6210 ± 0.0398 (5)	0.5774 ± 0.0641 (8)	0.6320 ± 0.0326 (4)
	(50)	($200, 2^{0}$)	($100, 2^{4}$)	($200, 2^{4}$)	($500, 2^{4}, 2^{4}$)
WDBC	0.6242 ± 0.0265 (9)	0.7477 ± 0.0243 (6)	0.7581 ± 0.0146 (5)	0.7318 ± 0.1425 (7)	0.7772 ± 0.0481 (4)
	(50)	($50, 2^{-6}$)	($100, 2^{0}$)	($100, 2^{-6}$)	($200, 2^{2}, 2^{2}$)
Australian	0.5843 ± 0.0310 (9)	0.7139 ± 0.0082 (7)	0.7446 ± 0.0326 (6)	0.7568 ± 0.0670 (4)	0.7874 ± 0.0747 (3)
	(500)	($50, 2^{-6}$)	($200, 2^{0}$)	($50, 2^{-2}$)	($500, 2^{6}, 2^{6}$)
German	0.5615 ± 0.0156 (9)	0.6086 ± 0.0120 (6)	0.5998 ± 0.0164 (8)	0.6177 ± 0.0379 (4)	0.6294 ± 0.0177 (3)
	(200)	($100, 2^{10}$)	($50, 2^{6}$)	($200, 2^{10}$)	($200, 2^{0}, 2^{0}$)
Ionosphere	0.6078 ± 0.0334 (9)	0.6882 ± 0.0172 (7)	0.6896 ± 0.0259 (6)	0.7116 ± 0.0549 (3)	0.6954 ± 0.0734 (4)
	(200)	($200, 2^{-6}$)	($200, 2^{10}$)	($100, 2^{-4}$)	($100, 2^{0}, 2^{0}$)
Heart	0.5705 ± 0.0146 (9)	0.6578 ± 0.0322 (8)	0.6936 ± 0.0099 (6)	0.6864 ± 0.1448 (7)	0.7147 ± 0.0622 (3)
	(50)	($100, 2^{-6}$)	($50, 2^{8}$)	($50, 2^{10}$)	($500, 2^{8}, 2^{8}$)
Madelon	0.5302 ± 0.0249 (9)	0.5615 ± 0.0189 (3)	0.5447 ± 0.0199 (5)	0.5546 ± 0.0144 (4)	0.5348 ± 0.0135 (7)
	(200)	($50, 2^{-6}$)	($500, 2^{10}$)	($200, 2^{-2}$)	($200, 2^{-2}, 2^{-2}$)
A3a	0.5935 ± 0.0357 (8)	0.6769 ± 0.0532 (2)	0.6128 ± 0.0421 (6)	0.6571 ± 0.0826 (4)	0.5969 ± 0.0551 (7)
	(50)	($100, 2^{-10}$)	($100, 2^{-4}$)	($100, 2^{-10}$)	($100, 2^{6}, 2^{6}$)
Splice	0.5781 ± 0.0311 (8)	0.6111 ± 0.0278 (5)	0.6167 ± 0.0337 (3)	0.6005 ± 0.0088 (7)	0.5757 ± 0.0411 (9)
	(200)	($50, 2^{-6}$)	($500, 2^{-6}$)	($50, 2^{2}$)	($50, 2^{2}, 2^{2}$)

Table 11 continued

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Liver-disorders	0.6741 ± 0.0099 (9)	0.7178 ± 0.0098 (7)	0.7198 ± 0.0116 (5)	0.7423 ± 0.1037 (3)	0.7411 ± 0.0223 (4)
	(500)	(500, 2^{-6})	(500, 2^{8})	(50, 2^{-10})	100, 2^{10}, 2^{10})
Breast-cancer	0.6873 ± 0.0353 (9)	0.8810 ± 0.0476 (6)	0.8943 ± 0.0195 (5)	0.9023 ± 0.0624 (4)	0.8785 ± 0.0293 (7)
	(50)	(50, 2^{2})	(50, 2^{-4})	(200, 2^{-8})	(100, 2^{-2}, 2^{-2})
Sonar	0.3363 ± 0.0315 (9)	0.4473 ± 0.0250 (5)	0.4009 ± 0.0408 (7)	0.4244 ± 0.1188 (6)	0.4686 ± 0.0357 (4)
	(50)	(100, 2^{-2})	(500, 2^{-2})	(50, 2^{2})	(50, 2^{2}, 2^{2})
Svmguide3	0.8332 ± 0.0088 (9)	0.9709 ± 0.0650 (7)	**1.0000 ± 0.0000 (3.5)**	**1.0000 ± 0.0000 (3.5)**	**1.0000 ± 0.0000 (3.5)**
	(50)	(50, 2^{-10})	(50, 2^{-10})	(50, 2^{-10})	(50, 2^{-10}, 2^{-10})
Colon-cancer	0.6892 ± 0.0602 (8)	0.7997 ± 0.0181 (6)	0.7905 ± 0.0489 (7)	0.8048 ± 0.0315 (5)	0.8239 ± 0.1237 (3)
	(100)	(50, 2^{2})	(100, 2^{-2})	(50, 2^{4})	(100, 2^{-10}, 2^{-10})
A1a	0.6031 ± 0.0507 (8)	0.6540 ± 0.0499 (3)	0.6123 ± 0.0246 (7)	0.6006 ± 0.0563 (9)	0.6235 ± 0.0307 (6)
	(100)	(200, 2^{10})	(200, 2^{-6})	(200, 2^{-6})	(100, 2^{2}, 2^{2})
Average rank	8.7333	5.6000	5.6333	5.2333	4.7667

Datasets	SPTELM (L, C_1, C_2, ε)	PIFTSVM ($C_1, C_2, \tau, \varepsilon$)	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM (L, C, τ, K, ς)
Diabetes	0.5999 ± 0.0541 (7)	0.6327 ± 0.0325 (3)	0.6362 ± 0.0616 (2)	**0.6683 ± 0.0540 (1)**
	(50, 2^{-8}, 2^{-8}, 0.4)	(2^{-10}, 2^{-10}, 0.25, 0)	(100, 2^{-2}, 0.25, 0.5)	(200, 2^{4}, 0.75, 13, 0.5)
WDBC	0.7050 ± 0.0458 (8)	0.8049 ± 0.0833 (3)	0.8116 ± 0.0803 (2)	**0.8791 ± 0.0500 (1)**
	(200, 2^{10}, 2^{10}, 0.4)	(2^{0}, 2^{0}, 0.25, 0)	(200, 2^{8}, 0.75, 0.5)	(50, 2^{0}, 0.75, 13, 0.5)

Table 11 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Australian	0.6413 ± 0.0338 (8)	0.7568 ± 0.0757 (5)	**0.8018 ± 0.0795 (1)**	0.7989 ± 0.0802 (2)
	$(100, 2^{-2}, 2^{-2}, 0.2)$	$(2^0, 2^0, 0.25, 0)$	$(50, 2^0, 0.25, 0.5)$	$(50, 2^0, 0.25, 7, 0.5)$
German	0.6027 ± 0.0482 (7)	0.6114 ± 0.0117 (5)	**0.6660 ± 0.0436 (1)**	0.6533 ± 0.0411 (2)
	$(100, 2^{-2}, 2^{-2}, 0.2)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(200, 2^6, 0.5, 0.5)$	$(50, 2^{10}, 2^{10}, 0.75, 17, 0.5)$
Ionosphere	0.6604 ± 0.0809 (8)	0.7220 ± 0.0924 (2)	0.6949 ± 0.1152 (5)	**0.7705 ± 0.0837 (1)**
	$(100, 2^2, 2^2, 0.4)$	$(2^0, 2^0, 0.25, 0)$	$(50, 2^8, 0.25, 0.5)$	$(200, 2^6, 2^6, 0.5, 15, 0.5)$
Heart	0.7007 ± 0.0404 (4)	0.7385 ± 0.0878 (2)	0.6995 ± 0.0692 (5)	**0.7681 ± 0.0825 (1)**
	$(200, 2^{-10}, 2^{-10}, 0.2)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(50, 2^4, 0.5, 0.5)$	$(100, 2^4, 0.5, 11, 15, 0.5)$
Madelon	0.5438 ± 0.0170 (6)	0.5326 ± 0.0134 (8)	0.5751 ± 0.0157 (2)	**0.5843 ± 0.0271 (1)**
	$(50, 2^{-2}, 2^{-2}, 0.4)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(200, 2^{-8}, 0.5, 0.5)$	$(50, 2^{-2}, 0.5, 1, 0.5)$
A3a	0.5730 ± 0.0141 (9)	0.6424 ± 0.0377 (5)	0.6574 ± 0.0702 (3)	**0.6946 ± 0.0825 (1)**
	$(50, 2^{-6}, 2^{-6}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{-10}, 0.25, 0.5)$	$(200, 2^2, 0.5, 17, 0.5)$
Splice	0.6117 ± 0.0236 (4)	0.6029 ± 0.0389 (6)	0.6475 ± 0.0295 (2)	**0.6624 ± 0.0435 (1)**
	$(50, 2^{-6}, 2^{-6}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{-6}, 0.5, 0.5)$	$(50, 2^2, 0.75, 9, 0.5)$
Liver-disorders	0.7184 ± 0.0217 (6)	0.7123 ± 0.0840 (8)	0.7439 ± 0.0319 (2)	**0.8010 ± 0.0158 (1)**
	$(50, 2^6, 2^6, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(50, 2^{-10}, 0.25, 0.5)$	$(50, 2^{-4}, 0.75, 9, 0.5)$
Breast-cancer	0.7711 ± 0.0300 (8)	0.9358 ± 0.0633 (2)	0.9172 ± 0.0477 (3)	**0.9605 ± 0.0264 (1)**
	$(50, 2^{-6}, 2^{-6}, 0.4)$	$(2^{-4}, 2^{-4}, 0.25, 0)$	$(50, 2^{-10}, 0.25, 0.5)$	$(50, 2^{-2}, 0.25, 5, 0.5)$

Table 11 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Sonar	0.4781 ± 0.0638 (3)	0.3841 ± 0.0799 (8)	0.4855 ± 0.1158 (2)	**0.5340 ± 0.0569 (1)**
	$(100, 2^{-10}, 2^{-10}, 0.2)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(100, 2^{4}, 0.25, 0.5)$	$(50, 2^{8}, 0.75, 7, 0.5)$
Svmguide3	0.8543 ± 0.0390 (8)	**1.0000 ± 0.0000 (3.5)**	**1.0000 ± 0.0000 (3.5)**	**1.0000 ± 0.0000 (3.5)**
	$(100, 2^{-6}, 2^{-6}, 0.2)$	$(2^{2}, 2^{2}, 0.25, 0)$	$(50, 2^{10}, 0.25, 0.5)$	$(50, 2^{-10}, 0.25, 1, 0.5)$
Colon-cancer	0.8091 ± 0.0512 (4)	0.6464 ± 0.2415 (9)	0.8912 ± 0.0450 (2)	**0.9513 ± 0.0059 (1)**
	$(200, 2^{-4}, 2^{-4}, 0.2)$	$(2^{10}, 2^{10}, 0.25, 0)$	$(50, 2^{8}, 0.25, 0.5)$	$(50, 2^{10}, 0.25, 3, 0.5)$
A1a	0.6316 ± 0.0322 (5)	0.7053 ± 0.0939 (4)	**0.7376 ± 0.0166 (1)**	0.7053 ± 0.0939 (2)
	$(100, 2^{2}, 2^{2}, 0)$	$(2^{-2}, 2^{-2}, 0.75, 0)$	$(200, 2^{2}, 0.75, 0.5)$	$(200, 2^{6}, 0.75, 11, 0.5)$
Average rank	6.3333	4.9000	2.4333	**1.3667**

The best results for each dataset are shown in bold

Table 12 F_1 of nine algorithms in the 50% label noise environment

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Diabetes	0.6533 ± 0.0143 (9)	0.7318 ± 0.0186 (7)	0.7774 ± 0.0352 (6)	0.7959 ± 0.0413 (3)	0.7950 ± 0.0209 (4)
	(50)	$(100, 2^{8})$	$(100, 2^{-2})$	$(100, 2^{10})$	$(100, 2^{-8}, 2^{-8})$
WDBC	0.5364 ± 0.0318 (9)	0.6425 ± 0.0640 (7)	0.6689 ± 0.0121 (6)	0.6753 ± 0.1266 (5)	0.7067 ± 0.0726 (4)
	(200)	$(50, 2^{-6})$	$(500, 2^{10})$	$(100, 2^{-6})$	$(200, 2^{2}, 2^{2})$
Australian	0.5284 ± 0.0288 (9)	0.6734 ± 0.0069 (7)	0.7121 ± 0.0386 (6)	0.7470 ± 0.0460 (4)	0.7715 ± 0.0570 (2)
	(500)	$(100, 2^{-6})$	$(200, 2^{0})$	$(50, 2^{-2})$	$(500, 2^{6}, 2^{6})$
German	0.4244 ± 0.0211 (9)	0.4737 ± 0.0192 (6)	0.4526 ± 0.0092 (8)	0.4997 ± 0.0446 (4)	0.5028 ± 0.0285 (3)
	(200)	$(500, 2^{10})$	$(100, 2^{8})$	$(50, 2^{-10})$	$(100, 2^{8}, 2^{8})$
Ionosphere	0.5231 ± 0.0304 (9)	0.6010 ± 0.0131 (6)	0.6157 ± 0.0301 (5)	0.6382 ± 0.0923 (3)	0.5965 ± 0.0508 (7)
	(200)	$(500, 2^{6})$	$(50, 2^{-10})$	$(100, 2^{-4})$	$(100, 2^{0}, 2^{0})$
Heart	0.5805 ± 0.0277 (9)	0.6795 ± 0.0391 (8)	0.7071 ± 0.0546 (7)	0.7505 ± 0.0777 (3)	0.7468 ± 0.0212 (5)
	(200)	$(500, 2^{-6})$	$(50, 2^{8})$	$(200, 2^{-6})$	$(500, 2^{-10}, 2^{-10})$
Madelon	0.5897 ± 0.0462 (8)	0.5834 ± 0.0579 (9)	0.6412 ± 0.0205 (6)	0.6675 ± 0.0092 (2)	0.6660 ± 0.0336 (5)
	(200)	$(100, 2^{6})$	$(500, 2^{-6})$	$(50, 2^{-2})$	$(50, 2^{-6}, 2^{-6})$
A3a	0.4094 ± 0.0410 (8)	0.5051 ± 0.0771 (2)	0.4291 ± 0.0621 (6)	0.4900 ± 0.1017 (3)	0.4248 ± 0.0463 (7)
	(50)	$(100, 2^{-10})$	$(100, 2^{-4})$	$(100, 2^{-10})$	$(50, 2^{2}, 2^{2})$
Splice	0.5693 ± 0.0497 (9)	0.6435 ± 0.0504 (8)	0.6482 ± 0.0714 (7)	0.6669 ± 0.0337 (5)	0.6828 ± 0.0063 (3)
	(100)	$(50, 2^{-10})$	$(500, 2^{-8})$	$(200, 2^{-6})$	$(50, 2^{-10}, 2^{-10})$

Table 12 continued

Datasets	ELM (L)	RELM (L, C)	OELM (L, C)	FELM (L, C)	TELM (L, C_1, C_2)
Liver-disorders	0.5426 ± 0.0108 (9)	0.6119 ± 0.0262 (5)	0.6000 ± 0.0033 (6)	0.5961 ± 0.0349 (7)	0.6279 ± 0.0641 (3)
	(500)	$(500, 2^{-10})$	$(500, 2^{8})$	$(50, 2^{-10})$	$(200, 2^{-6}, 2^{-6})$
Breast-cancer	0.7138 ± 0.0383 (9)	0.9089 ± 0.0306 (6)	0.9233 ± 0.0183 (5)	0.9268 ± 0.0319 (4)	0.9051 ± 0.0271 (7)
	(50)	$(50, 2^{-6})$	$(50, 2^{-4})$	$(200, 2^{-8})$	$(100, 2^{-2}, 2^{-2})$
Sonar	0.5644 ± 0.0345 (9)	0.6705 ± 0.0224 (6)	0.6655 ± 0.0227 (7)	0.6595 ± 0.0979 (8)	0.7002 ± 0.0702 (2)
	(50)	$(100, 2^{10})$	$(500, 2^{-2})$	$(50, 2^{2})$	$(200, 2^{4}, 2^{4})$
Svmguide3	0.4096 ± 0.0252 (8)	0.4279 ± 0.0815 (7)	**0.4848 ± 0.0565 (1)**	0.3993 ± 0.0863 (9)	0.4526 ± 0.0554 (6)
	(50)	$(200, 2^{6})$	$(200, 2^{10})$	$(100, 2^{10})$	$(100, 2^{-2}, 2^{-2})$
Colon-cancer	0.5779 ± 0.0311 (8)	0.6953 ± 0.0638 (7)	0.7193 ± 0.0232 (5)	0.7187 ± 0.0223 (6)	0.7414 ± 0.0732 (3)
	(100)	$(50, 2^{2})$	$(50, 2^{2})$	$(50, 2^{-10})$	$(100, 2^{-10}, 2^{-10})$
A1a	0.4109 ± 0.0396 (9)	0.4665 ± 0.0505 (3)	0.4176 ± 0.0276 (7)	0.4147 ± 0.0364 (8)	0.4404 ± 0.0188 (6)
	(100)	$(200, 2^{10})$	$(200, 2^{-6})$	$(200, 2^{-6})$	$(50, 2^{6}, 2^{6})$
Average rank	8.7333	6.2667	5.8667	4.9333	4.4667

Datasets	SPTELM (L, C_1, C_2, ε)	PIFTSVM ($C_1, C_2, \tau, \varepsilon$)	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM (L, C, τ, K, ς)
Diabetes	0.6916 ± 0.0407 (8)	0.8011 ± 0.0236 (2)	0.7779 ± 0.0463 (5)	**0.8110 ± 0.0247 (1)**
	$(50, 2^{-8}, 2^{-8}, 0.4)$	$(2^{0}, 2^{0}, 0.25, 0)$	$(50, 2^{-2}, 0.75, 0.5)$	$(200, 2^{8}, 0.5, 5, 0.5)$
WDBC	0.6044 ± 0.0583 (8)	0.7532 ± 0.0918 (2)	0.7367 ± 0.1063 (3)	**0.8077 ± 0.1037 (1)**
	$(200, 2^{10}, 2^{10}, 0.4)$	$(2^{2}, 2^{2}, 0.75, 0)$	$(200, 2^{10}, 0.75, 0.5)$	$(50, 2^{-2}, 0.25, 11, 0.5)$

Table 12 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Australian	0.6051 ± 0.0580 (8)	0.7202 ± 0.7202 (5)	0.7604 ± 0.1056 (3)	**0.7724 ± 0.1057 (1)**
	$(100, 2^{-2}, 2^{-2}, 0.2)$	$(2^0, 2^0, 0.25, 0)$	$(500, 2^0, 0.75, 0.5)$	$(50, 2^0, 0.25, 7, 0.5)$
German	0.4757 ± 0.0441 (5)	0.4730 ± 0.0598 (7)	**0.5356 ± 0.0688 (1)**	0.5155 ± 0.0657 (2)
	$(100, 2^2, 2^2, 0.2)$	$(2^2, 2^2, 0.25, 0)$	$(200, 2^6, 0.5, 0.5)$	$(50, 2^{10}, 0.75, 17, 0.5)$
Ionosphere	0.5956 ± 0.0585 (8)	0.6490 ± 0.0734 (2)	0.6233 ± 0.1473 (4)	**0.6988 ± 0.1012 (1)**
	$(100, 2^2, 2^2, 0.4)$	$(2^2, 2^2, 0.75, 0)$	$(50, 2^8, 0.25, 0.5)$	$(200, 2^6, 0.5, 15, 0.5)$
Heart	0.7250 ± 0.0429 (6)	0.7841 ± 0.0252 (2)	0.7498 ± 0.0752 (4)	**0.8052 ± 0.0836 (1)**
	$(200, 2^{-10}, 2^{-10}, 0.2)$	$(2^{-2}, 2^{-2}, 0.25, 0)$	$(200, 2^2, 0.25, 0.5)$	$(100, 2^4, 0.5, 11, 0.5)$
Madelon	0.6673 ± 0.0208 (3)	0.6662 ± 0.0357 (4)	0.6176 ± 0.0731 (7)	**0.6726 ± 0.0223 (1)**
	$(100, 2^0, 2^0, 0.4)$	$(2^6, 2^6, 0.75, 0)$	$(50, 2^{-2}, 0.25, 0.5)$	$(50, 2^{10}, 0.5, 5, 0.5)$
A3a	0.3918 ± 0.0401 (9)	0.4676 ± 0.0410 (5)	0.4755 ± 0.0573 (4)	**0.5308 ± 0.0711 (1)**
	$(50, 2^{-6}, 2^{-6}, 0)$	$(2^{-10}, 2^{-10}, 0.25, 0)$	$(100, 2^{-2}, 0.5, 0.5)$	$(200, 2^{-6}, 0.5, 15, 0.5)$
Splice	0.6598 ± 0.0197 (6)	0.6833 ± 0.0134 (2)	0.6770 ± 0.0468 (4)	**0.7250 ± 0.0228 (1)**
	$(50, 2^6, 2^6, 0)$	$(2^2, 2^2, 0.25, 0)$	$(50, 2^{-10}, 0.5, 0.5)$	$(50, 2^{-6}, 0.5, 15, 0.5)$
Liver-disorders	0.6145 ± 0.0298 (4)	0.6290 ± 0.0712 (2)	0.5948 ± 0.0140 (8)	**0.6781 ± 0.0520 (1)**
	$(50, 2^6, 2^6, 0)$	$(2^4, 2^4, 0.75, 0)$	$(50, 2^{10}, 0.25, 0.5)$	$(500, 2^4, 0.5, 1, 0.5)$
Breast-cancer	0.7870 ± 0.0510 (8)	0.9497 ± 0.0376 (2)	0.9352 ± 0.0331 (3)	**0.9557 ± 0.0394 (1)**
	$(50, 2^2, 2^2, 0.2)$	$(2^{-4}, 2^{-4}, 0.25, 0)$	$(100, 2^{10}, 0.75, 0.5)$	$(50, 2^{-2}, 0.25, 5, 0.5)$

Table 12 continued

Datasets	SPTELM $(L, C_1, C_2, \varepsilon)$	PIFTSVM $(C_1, C_2, \tau, \varepsilon)$	TPin-IFELM-SFA (L, C, τ, ς)	TPin-IFELM $(L, C, \tau, K, \varsigma)$
Sonar	0.6718 ± 0.0824 (5)	0.6734 ± 0.0922 (4)	0.6972 ± 0.0751 (3)	$\mathbf{0.7255 \pm 0.0512}$ **(1)**
	$(100, 2^{-10}, 2^{-10}, 0.2)$	$(2^4, 2^4, 0.75, 0)$	$(100, 2^{-8}, 0.5, 0.5)$	$(50, 2^8, 0.75, 7, 0.5)$
Svmguide3	0.4610 ± 0.0583 (4)	0.4578 ± 0.0454 (5)	0.4628 ± 0.0925 (3)	0.4741 ± 0.0607 (2)
	$(100, 2^6, 2^6, 0.2)$	$(2^{-6}, 2^{-6}, 0.25, 0)$	$(50, 2^{-2}, 0.5, 0.5)$	$(50, 2^6, 0.5, 3, 0.5)$
Colon-cancer	0.7370 ± 0.1469 (4)	0.5218 ± 0.2657 (9)	0.8046 ± 0.0921 (2)	$\mathbf{0.8681 \pm 0.0182}$ **(1)**
	$(200, 2^{-4}, 2^{-4}, 0.2)$	$(2^{10}, 2^{10}, 0.25, 0)$	$(50, 2^8, 0.25, 0.5)$	$(50, 2^2, 0.75, 9, 0.5)$
A1a	0.4505 ± 0.0465 (5)	0.4593 ± 0.0378 (4)	$\mathbf{0.5598 \pm 0.0286}$ **(1)**	0.5340 ± 0.0760 (2)
	$(50, 2^{-2}, 2^{-2}, 0.4)$	$(2^{-2}, 2^{-2}, 0.75, 0)$	$(50, 2^{-10}, 0.25, 0.5)$	$(200, 2^{10}, 0.5, 9, 0.5)$
Average rank	6.0667	3.8000	3.6667	**1.2000**

The best results for each dataset are shown in bold

References

1. Huang G, Huang G, Song S, You K (2015) Trends in extreme learning machines: a review. Neural Netw 61:32–48
2. Huang G, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Cybernet 42:513–529
3. Huang G, Ding X, Zhou H (2010) Optimization method based extreme learning machine for classification. Neurocomputing 74(1–3):155–163
4. Sun P, Yang L (2022) Generalized eigenvalue extreme learning machine for classification. Appl Intell 52(6):6662–6691
5. Ahuja B, Vishwakarma VP (2021) Deterministic multi-kernel based extreme learning machine for pattern classification. Expert Syst Appl 183:115308
6. Ren L, Liu J, Gao Y, Kong X, Zheng C (2021) Kernel risk-sensitive loss based hyper-graph regularized robust extreme learning machine and its semi-supervised extension for classification. Knowl-Based Syst 227:107226
7. Wong H, Leung H, Leung C, Wong E (2022) Noise/fault aware regularization for incremental learning in extreme learning machines. Neurocomputing 486:200–214
8. Luo J, Wong C, Vong C (2021) Multinomial bayesian extreme learning machine for sparse and accurate classification model. Neurocomputing 423:24–33
9. Liu Z, Jin W, Mu Y (2020) Variances-constrained weighted extreme learning machine for imbalanced classification. Neurocomputing 403:45–52
10. Zong W, Huang G, Chen Y (2013) Weighted extreme learning machine for imbalance learning. Neurocomputing 101:229–242
11. Li Y, Zhang J, Zhang S, Xiao W, Zhang Z (2022) Multi-objective optimization-based adaptive class-specific cost extreme learning machine for imbalanced classification. Neurocomputing 496:107–120
12. Xiao W, Zhang J, Li Y, Zhang S, Yang W (2017) Class-specific cost regulation extreme learning machine for imbalanced classification. Neurocomputing 261:70–82
13. Dutta AK, Qureshi B, Albagory Y, Alsanea M, Al Faraj M, Sait ARW (2023) Optimal weighted extreme learning machine for cybersecurity fake news classification. Comput Syst Sci Eng 44(3):2395–2409
14. Tummalapalli S, Kumar L, Neti LBM, Krishna A (2022) Detection of web service anti-patterns using weighted extreme learning machine. Comput Stand Interfaces 82:103621
15. El Bourakadi D, Yahyaouy A, Boumhidi J (2022) Improved extreme learning machine with autoencoder and particle swarm optimization for short-term wind power prediction. Neural Comput Appl 34(6):4643–4659
16. Xia J, Yang D, Zhou H, Chen Y, Zhang H, Liu T, Heidari AA, Chen H, Pan Z (2022) Evolving kernel extreme learning machine for medical diagnosis via a disperse foraging sine cosine algorithm. Comput Biol Med 141:105137
17. Lin Z, Gao Z, Ji H, Zhai R, Shen X, Mei T (2022) Classification of cervical cells leveraging simultaneous super-resolution and ordinal regression. Appl Soft Comput 115:108208
18. Gao Z, Hu Q, Xu X (2022) Condition monitoring and life prediction of the turning tool based on extreme learning machine and transfer learning. Neural Comput Appl 34(5):3399–3410
19. Wang Y, Li R, Chen Y (2021) Accurate elemental analysis of alloy samples with high repetition rate laser-ablation spark-induced breakdown spectroscopy coupled with particle swarm optimization-extreme learning machine. Spectrochim Acta Part B-Atomic Spectrosc 177:106077
20. Wu D, Wang X, Wu S (2022) A hybrid framework based on extreme learning machine, discrete wavelet transform, and autoencoder with feature penalty for stock prediction. Expert Syst Appl 207:118006
21. Wang GC, Zhang Q, Band SS, Dehghani M, Chau KW, Tho QT, Zhu S, Samadianfard S, Mosavi A (2022) Monthly and seasonal hydrological drought forecasting using multiple extreme learning machine models. Eng Appl Comput Fluid Mech 16(1):1364–1381
22. Wang L, Khishe M, Mohammadi M, Mahmoodzadeh A (2022) Extreme learning machine evolved by fuzzified hunger games search for energy and individual thermal comfort optimization. J Build Eng 60:105187
23. Al-Yaseen WL, Idrees AK, Almasoudy FH (2022) Wrapper feature selection method based differential evolution and extreme learning machine for intrusion detection system. Pattern Recogn 132:108912
24. Ren Z, Yang L (2018) Correntropy-based robust extreme learning machine for classification. Neurocomputing 313:74–84
25. Wang Y, Yang L, Yuan C (2019) A robust outlier control framework for classification designed with family of homotopy loss function. Neural Netw 112:41–53
26. Ren Z, Yang L (2019) Robust extreme learning machines with different loss functions. Neural Process Lett 49(3):1543–1565

27. Shen J, Ma J (2019) Sparse twin extreme learning machine with epsilon-insensitive zone pinball loss. IEEE Access 7:112067–112078
28. Huang Z, Li J (2022) Discernibility measures for fuzzy β covering and their application. IEEE Trans Cybernet 52(9):9722–9735
29. Huang Z, Li J, Qian Y (2022) Noise-tolerant fuzzy-β-covering-based multigranulation rough sets and feature subset selection. IEEE Trans Fuzzy Syst 30(7):2721–2735
30. Huang, Z., Li, J.: Noise-tolerant discrimination indexes for fuzzy γ covering and feature subset selection. IEEE Trans Neural Netw Learn Syst (Early Access)
31. Lin C, Wang S (2002) Fuzzy support vector machines. IEEE Trans Neural Netw 13(2):464–471
32. Zhang W, Ji H (2013) Fuzzy extreme learning machine for classification. Electron Lett 49(7):448–449
33. Shen X, Niu L, Qi Z, Tian Y (2017) Support vector machine classifier with truncated pinball loss. Pattern Recogn 68:199–210
34. Wang H, Xu Y, Zhou Z (2021) Twin-parametric margin support vector machine with truncated pinball loss. Neural Comput Appl 33(8):3781–3798
35. Yuille A, Rangarajan A (2003) The concave-convex procedure. Neural Comput 15(4):915–936
36. Lipp T, Boyd S (2016) Variations and extension of the convex-concave procedure. Optim Eng 17(2):263–287
37. Huang G, Zhu Q, Siew C-K (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
38. Rezvani S, Wang X, Pourpanah F (2019) Intuitionistic fuzzy twin support vector machines. IEEE Trans Fuzzy Syst 27(11):2140–2151
39. Liang Z, Zhang L (2022) Intuitionistic fuzzy twin support vector machines with the insensitive pinball loss. Appl Soft Comput 115:108231
40. Laxmi S, Gupta SK (2022) Multi-category intuitionistic fuzzy twin support vector machines with an application to plant leaf recognition. Eng Appl Artif Intell 110:104687
41. Wong CM, Vong CM, Wong PK, Cao J (2018) Kernel-based multilayer extreme learning machines for representation learning. IEEE Trans Neural Netw Learn Syst 29(3):757–762
42. Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30