Steel Plate Surface Defects Classification Method using Multiple Hyper-planes Twin Support Vector Machine with Additional Information

Chu Maoxiang, Zhai Zixuan, Liu Liming*, Liu Guanghu

***Abstract*—Focusing on the actual situation of steel surface defects, a novel multiple hyper-planes twin support vector machine with additional information (MHTSVM) is proposed. Similar to twin support vector machine (TSVM), MHTSVM also uses the nonparallel hyper-planes to deal with the classification problem. Differently, the MHTSVM model is trained with expert samples and rough samples respectively, which makes the MHTSVM more applicable. In addition, MHTSVM introduces the gaussian weight information obtained with K-nearest neighbors (KNN) method, which makes MHTSVM deal with noise samples and reliable. Meanwhile, the within-class scatter information is embedded into MHTSVM model. The scatter information can be used to balance the samples distribution in defect dataset. The above two kinds of additional information make the MHTSVM have perfect classification performance in accuracy. Lastly, the MHTSVM algorithm is used to classify six types of steel surface defects. And a series of numerical experiments are performed. The experimental results proved that the novel MHTSVM model has perfect ability, especially for corrupted defect samples.**

***Index Terms*—steel surface defects, multi-class classification, twin support vector machine, additional information**

I. Introduction

In recent years, the iron and steel industry in China has made great improvements in production equipment and product quality. Though the annual output often ranks first in the world, it still cannot face the severe market competition and high-quality requirements freely. The surface quality of steel plate is one of the important standards to measure whether the steel plate is qualified or not. Among them, surface defects are the most important problem that hinders iron and steel enterprises from improving the quality of steel products. Therefore, the research on the steel surface defect detection technology has become the key. The defect detection system [1-2] usually includes three sub-systems. They are defect acquisition sub-system, defect location sub-system and defect classification sub-system. Among them, the defects classification is the most important.

The accuracy of defect classification directly affects the detection result. Accurate defect classification model plays an important role in ensuring steel product quality, improving production efficiency and reducing cost. Due to the problems of high labor intensity, low efficiency and error prone in manual classification, the development of classification model based on pattern recognition has attracted extensive attention in recent decades. Many researchers [3-9] have studied the strip steel surface defects classification with pattern recognition algorithms, such as K-nearest neighbors (KNN) method and decision tree method. Methods mentioned above are based on statistical learning theory [10]. Each method has unique advantages to deal with different classification problems. Neogi [11] made a comprehensive review for the current defect detection and classification algorithms based on vision technology. According to his research, we found that neural network (NN) [12] and support vector machine (SVM) [13] are the most commonly used classifiers. However, the classifiers based on NN model usually get into the local minimum and need a large number of training samples. On the contrary, SVM can avoid the problems of NN algorithm. On the other hand, SVM is better than NN algorithm in high-dimensional problems. SVM has better generalization ability. Therefore, the steel surface defects classification model based on SVM has a very broad application prospect in iron and steel enterprises.

Besides the above traditional classification algorithms, the most popular pattern recognition algorithm is deep learning [14], which is an effective feature learning method with strong learning ability. Deep learning has been applied in the mainstream field of computer vision, but in the industrial field, such as product surface defect detection, the powerful ability of deep learning has not been brought into full play. This is mainly because of difficulties in industry data collection and long-term accumulation. The defect data sets collected from different steel production lines are small and medium-sized. Faced with such data sets, deep learning is very limited and incompetent because it needs a large number of training samples. On the contrary, the classification algorithm based on SVM has proved to be one of the most effective methods.

Manuscript received September 23, 2022; revised June 2, 2023.

This work was supported in part by the Natural Science Foundation of Liaoning Province of China (2022-MS-353), Liaoning Province Ministry of Education Scientific Study Project (2020LNZD06).

M. X. Chu is a professor and doctoral supervisor in University of Science and Technology Liaoning, China. (e-mail: chu52_2004@163.com).

Z. X. Zhai is a postgraduate student in University of Science and Technology Liaoning, China. (e-mail: zzx15842061275@163.com).

L. M. Liu is a PhD candidate of University of Science and Technology Liaoning, China. (Corresponding author to provide phone: 18341252046; e-mail: llm06101021@hotmail.com).

G. H. Liu is a PhD candidate of University of Science and Technology Liaoning, China. (e-mail: lqgdhu@163.com).

Although the support vector machine algorithm realizes the defect classification of different steels, it does not meet the actual situation of steel surface defects completely. In the actual situation, there are a few expert samples and a large number of rough samples. The so-called expert sample is the standard sample specified by experts, which has a certain authority. They are common on site, but the number is relatively small. The rough sample is extracted from the production line by front-line workers with many years of practical experience. Although the number of rough samples is large, they are probably wrong samples in label. The existence of expert samples and rough samples is consistent with the actual industrial production. Thus, it is necessary to reconstruct the classification model with the expert samples and rough samples. Therefore, in this paper, based on twin support vector machine (TSVM) [15], we propose a steel plate surface defect classification model called multiple hyper-planes TSVM with additional information (MHTSVM). MHTSVM has the following properties and advantages.

(1) For the expert samples and rough samples, a novel MHTSVM model is developed. In the training stage, the MHTSVM model is obtained by learning expert samples and rough samples, respectively, so that the classification model is more in line with the actual industrial production.

(2) The new MHTSVM is based on TSVM and extended by using the additional information mined with two extraction methods [16-17]. On one hand, MHTSVM formulates TSVM classification model, which makes MHTSVM have good generalization performance. On the other hand, there are two kinds of additional information in MHTSVM algorithm: gaussian weight information and within-class scatter information. They are obtained by extracting the classification information in the original training samples. The additional information not only suppresses the adverse impact of label noise, but also considers the difference of samples size and distribution. Above all, the classification model based on MHTSVM has stronger applicability.

(3) The MHTSVM is used to classify six types of steel surface defects. The experimental results show that our proposed MHTSVM model has higher accuracy and stronger robustness, especially for the defect samples with noise.

II. Defect Samples Set for Steel Surface

There are many types of defects on strip steel surface. Combined with the on-line production line of hot rolling mill, the defects can be divided into two parts. One part is the defect samples given by technologists, which is called expert samples. These samples are very common on steel production lines. Based on personal experience, technologists define the defect types according to the defect shape and feature. For example, the black spots including spot, stripe and scale on the surface of steel plate are called rolled-in-scale; the scaly and striped metal layers are called scarring. Because the expert samples are given manually, they are characterized by small sample size, but they are completely correct. The other part of the defect samples is directly collected from on-line production line. The number of these samples is much higher than that of the expert samples. This kind of defect samples is named as rough samples. Labels of these samples are defined by production line workers. Some samples maybe misclassified due to the workers' inadequate experience. To sum up, rough samples are characterized by large sample size but noise.

The existence of expert database and rough database accords with the actual industrial production. In this paper, six types of defect images are collected, such as scratches (SH), inclusion (IN), patches (PH), rolled-in-scale (RD), pitted-surface (PE), and crazing (CZ). The typical defect images are shown in Fig. 1. The total of defect images is 3516, including 155 defect images in expert database, 1211 defect images in rough database and 2150 defect images in testing database. In order to yield defects dataset for training and testing, those defect images are processed in advance [18-20]. Then, each defect region is described with a 52-dimensional feature vector. These feature vectors make up the defect samples set. Lastly, the defect samples set is split into training and testing samples as Table I. It can be easily seen that the expert samples (ExSa) are much less than the rough samples (RoSa). (there are few training expert samples (ExSa) and relatively many training rough samples (RoSa)).

TABLE I
ATTRIBUTE OF STEEL SURFACE DEFECTS DATASET

Defect type	Total samples	Training ExSa	Training RoSa	Testing samples
CZ	556	25	177	354
IN	815	30	262	523
PH	710	25	229	456
PE	350	25	108	217
RD	642	25	296	321
SH	443	25	139	279
Total	**3516**	**155**	**1211**	**2150**

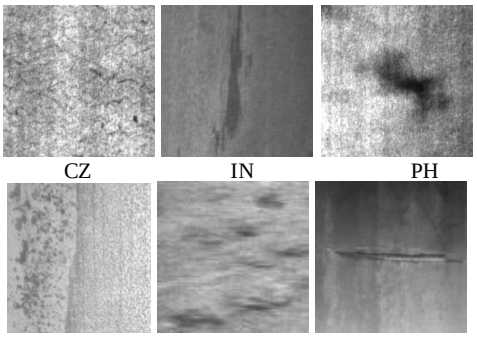


Fig. 1. Six types of steel surface defect images.

III. Multiple Hyper-planes Twin Support Vector Machine with Additional Information

A. Model of MHTSVM classifier

Suppose the defect samples set is $\boldsymbol{X}=[\bar{\boldsymbol{X}}^T\ \hat{\boldsymbol{X}}^T]^T\in\Re^{m\times d}$, where m represents the total number of defect samples, and d is the dimension of the feature space. $\bar{\boldsymbol{X}}=\{\bar{\boldsymbol{X}}^k\in\Re^{\bar{m}^k\times d}\,|\,k=1,.....,N\}$ denotes the expert defects database, and $\hat{\boldsymbol{X}}=\{\hat{\boldsymbol{X}}^k\in\Re^{\hat{m}^k\times d}\,|\,k=1,.....,N\}$ denotes the rough defects database, where $\bar{\boldsymbol{X}}^k$ or $\hat{\boldsymbol{X}}^k$ is the kth defect samples subset, $\bar{m}^k$ or $\hat{m}^k$ represents the number of $\bar{\boldsymbol{X}}^k$ or $\hat{\boldsymbol{X}}^k$, and N refers to the types of defects. In this paper, we have $N=6$ and $d=52$.

Under the existence of expert samples and rough samples, a multi-class classification model is established based on MHTSVM, which is used to realize the surface defects classification in this paper. The multi-class classification model is made up of two phases: training phase and testing phase. In the training phase, the binary tree method is employed, and $N-1$ sub-models called as MHTSVM-*k* are built. For binary classification, two nonparallel hyper-planes of MHTSVM-*k* are constructed just like TSVM.

$$h_k^+: \ \varphi^T(\boldsymbol{x})\boldsymbol{w}_k^+ + b_k^+ = 0 \quad \text{and} \quad h_k^-: \ \varphi^T(\boldsymbol{x})\boldsymbol{w}_k^- + b_k^- = 0, \tag{1}$$

where $\varphi(\cdot)$ is a feature space mapping function which maps features to a higher or infinite dimensional space, $\boldsymbol{w}_k^+$, $\boldsymbol{w}_k^-$ are weight vectors, b_k^+ and b_k^- are the biases. Once the augmented vectors $[\boldsymbol{w}_1^T, b_1]^T$ and $[\boldsymbol{w}_2^T, b_2]^T$ are calculated, the new sample $\boldsymbol{x}_t$ can be classified. As for N types of defects, a new sample $\boldsymbol{x}_t$ is firstly calculated in each $[(\boldsymbol{w}_k^+)^T, b_k^+]^T$ and $[(\boldsymbol{w}_k^-)^T, b_k^-]^T$, where $k = 1, 2, ..., N-1$. Then, the final classification result is output through the prediction function $f_t(\boldsymbol{x}_t, \{[(\boldsymbol{w}_k^+)^T, b_k^+]^T, [(\boldsymbol{w}_k^-)^T, b_k^-]^T\}_{k=1}^{N-1})$.

MHTSVM-*k* has the structure of nonparallel hyper-planes in TSVM. What's more, MHTSVM-*k* not only uses the expert samples, but also uses the rough samples. Therefore, MHTSVM-*k* can improve the classification accuracy. Meanwhile, the gaussian weight information is introduced into MHTSVM-*k*, which reduce the sensibility of TSVM in rough samples with label noise. In addition, the unbalance of samples size and distribution is considered simultaneously in MHTSVM-*k*. It can solve the problem caused by unbalanced data in TSVM. To sum up, MHTSVM has perfect accuracy and strong robustness for steel surface defects dataset, especially for corrupted defects dataset.

B. MHTSVM-k formula

For *N* kinds of defects, MHTSVM model only requires to construct *N*-1 MHTSVM-*k* sub-models. For each MHTSVM-*k* sub-model, two TSVM-type QPPs are solved and two independent classification hyper-planes are obtained. Differently, the MHTSVM-*k* sub-model makes full use of gaussian weight information and within-class scatter information for surface defect samples, which improves the anti-noise ability and classification accuracy. The QPPs of MHTSVM-*k* can be expressed as follows.

$$\begin{aligned}
\min_{\boldsymbol{w}_1^k, b_1^k} \ & \frac{\bar{S}^k}{2\bar{m}^k}\sum_{\bar{i}=1}^{\bar{m}^k} K\left(\varphi(\bar{\boldsymbol{x}}_{\bar{i}}^k)^T\boldsymbol{w}_k^+ + b_k^+\right)^2 \\
& + \frac{\hat{S}^k}{2\hat{m}^k}\sum_{\hat{i}=1}^{\hat{m}^k}\hat{\rho}_{\hat{i}}^k\left(\varphi(\hat{\boldsymbol{x}}_{\hat{i}}^k)^T\boldsymbol{w}_k^+ + b_k^+\right)^2 \\
& + \frac{\bar{C}_1^k\bar{\Upsilon}^k}{\bar{m}^r}\sum_{o=k+1}^{N}\bar{T}^{k,o}\sum_{\bar{j}=1}^{\bar{m}^o}K\bar{\eta}_{\bar{j}} + \frac{\hat{C}_1^k\hat{\Upsilon}^k}{\hat{m}^r}\sum_{o=k+1}^{N}\hat{T}^{k,o}\sum_{\hat{j}=1}^{\hat{m}^o}\hat{\rho}_{\hat{j}}^o\hat{\eta}_{\hat{j}} \\
\text{s.t.} \ & \varphi(\bar{\boldsymbol{x}}_{\bar{j}}^o)^T\boldsymbol{w}_k^+ + b_k^+ \le -1 + \bar{\eta}_{\bar{j}}, \\
& \varphi(\hat{\boldsymbol{x}}_{\hat{j}}^o)^T\boldsymbol{w}_k^+ + b_k^+ \le -1 + \hat{\eta}_{\hat{j}}, o = k+1, \cdots, N, \\
& \bar{\eta}_{\bar{j}} \ge 0, \hat{\eta}_{\hat{j}} \ge 0, \bar{j} = 1, \cdots, \bar{m}^o, \hat{j} = 1, \cdots, \hat{m}^o,
\end{aligned} \tag{2}$$

$$\begin{aligned}
\min_{\boldsymbol{w}_2^k, b_2^k} \ & \frac{\bar{\Upsilon}^k}{2\bar{m}^r}\sum_{o=k+1}^{N}\bar{T}^{k,o}\sum_{\bar{j}=1}^{\bar{m}^o}K\left(\varphi(\bar{\boldsymbol{x}}_{\bar{j}}^o)^T\boldsymbol{w}_k^- + b_k^-\right)^2 \\
& + \frac{\hat{\Upsilon}^k}{2\hat{m}^r}\sum_{o=k+1}^{N}\hat{T}^{k,o}\sum_{\hat{j}=1}^{\hat{m}^o}\hat{\rho}_{\hat{j}}^o\left(\varphi(\bar{\boldsymbol{x}}_{\hat{j}}^o)^T\boldsymbol{w}_k^- + b_k^-\right)^2 \\
& + \frac{\bar{C}_2^k\bar{S}^k}{\bar{m}^k}\sum_{\bar{i}=1}^{\bar{m}^k}K\bar{\xi}_{\bar{i}} + \frac{\hat{C}_2^k\hat{S}^k}{\hat{m}^k}\sum_{\hat{i}=1}^{\hat{m}^k}\hat{\rho}_{\hat{i}}^k\hat{\xi}_{\hat{i}} \\
\text{s.t.} \ & \varphi(\bar{\boldsymbol{x}}_{\bar{i}}^k)^T\boldsymbol{w}_k^- + b_k^- \ge 1 - \bar{\xi}_{\bar{i}}, \\
& \varphi(\hat{\boldsymbol{x}}_{\hat{i}}^k)^T\boldsymbol{w}_k^- + b_k^- \ge 1 - \hat{\xi}_{\hat{i}}, \\
& \bar{\xi}_{\bar{i}} \ge 0, \hat{\xi}_{\hat{i}} \ge 0, \bar{i} = 1, \cdots, \bar{m}^k, \hat{i} = 1, \cdots, \hat{m}^k,
\end{aligned} \tag{3}$$

where $\bar{m}^r = \sum_{o=k+1}^{N}\bar{m}^o$ and $\hat{m}^r = \sum_{o=k+1}^{N}\hat{m}^o$ can be determined, $\bar{C}_1^k$, $\hat{C}_1^k$, $\bar{C}_2^k$ and $\hat{C}_2^k$ are positive penalty parameters, $\bar{\eta}_{\bar{j}}$, $\hat{\eta}_{\hat{j}}$, $\bar{\xi}_{\bar{i}}$ and $\hat{\xi}_{\hat{i}}$ are slack variables. $\hat{\rho}_{\hat{i}}^k$ and $\hat{\rho}_{\hat{j}}^o$ are gaussian weight, $\bar{S}^k$, $\hat{S}^k$, $\bar{\Upsilon}^k$, $\hat{\Upsilon}^k$, $\bar{T}^{k,o}$ and $\hat{T}^{k,o}$ are the within-class scatter. The computational process of gaussian weight information and the within-class scatter is described in previous work [16-17].

Take the QPP (2) as an example, the illustrations are described. Firstly, as can be seen that there are four terms in objective function. Unlike TSVM algorithm, the expert samples and rough samples are considered respectively. This makes our MHTSVM model more consistent with the actual production line. Secondly, the gaussian weight information is introduced into each term. This contribution makes the noise samples be weakened and the important samples are strengthened. Thirdly, MHTSVM-*k* not only uses the different penalty parameters $\bar{C}_1^k$ and $\hat{C}_1^k$, but also divides each term by the number of samples, which can effectively cope with the problem of unbalanced samples size. Furthermore, the within-class scatter information is employed. It ensures that MHTSVM-*k* can handle the problem of unbalanced distribution between two classes.

C. Solution to MHTSVM-k

In order to solve the QPP (2) for MHTSVM-*k* sub-model, the Lagrange function is first constructed with the following formula:

$$\begin{aligned}
L = & \frac{\bar{S}^k}{2\bar{m}^k}\sum_{\bar{i}=1}^{\bar{m}^k}K\left(\varphi(\bar{\boldsymbol{x}}_{\bar{i}}^k)^T\boldsymbol{w}_k^+ + b_k^+\right)^2 - \sum_{o=k+1}^{N}\sum_{\bar{j}=1}^{\bar{m}^o}\bar{\beta}_{\bar{j}}^o\bar{\eta}_{\bar{j}}^o \\
& + \frac{\hat{S}^k}{2\hat{m}^k}\sum_{\hat{i}=1}^{\hat{m}^k}\hat{\rho}_{\hat{i}}^k\left(\varphi(\hat{\boldsymbol{x}}_{\hat{i}}^k)^T\boldsymbol{w}_k^+ + b_k^+\right)^2 - \sum_{o=k+1}^{N}\sum_{\hat{j}=1}^{\hat{m}^o}\hat{\beta}_{\hat{j}}^o\hat{\eta}_{\hat{j}}^o \\
& + \frac{\bar{C}_1^k\bar{\Upsilon}^k}{\bar{m}^r}\sum_{o=k+1}^{N}\bar{T}^{k,o}\sum_{\bar{j}=1}^{\bar{m}^o}K\bar{\eta}_{\bar{j}} + \frac{\hat{C}_1^k\hat{\Upsilon}^k}{\hat{m}^r}\sum_{o=k+1}^{N}\hat{T}^{k,o}\sum_{\hat{j}=1}^{\hat{m}^o}\hat{\rho}_{\hat{j}}^o\hat{\eta}_{\hat{j}} \\
& + \sum_{o=k+1}^{N}\sum_{\bar{j}=1}^{\bar{m}^o}\bar{\alpha}_{\bar{j}}^o\left(\varphi(\bar{\boldsymbol{x}}_{\bar{j}}^o)^T\boldsymbol{w}_k^+ + b_k^+ + 1 - \bar{\eta}_{\bar{j}}\right) \\
& + \sum_{o=k+1}^{N}\sum_{\hat{j}=1}^{\hat{m}^o}\hat{\alpha}_{\hat{j}}^o\left(\varphi(\hat{\boldsymbol{x}}_{\hat{j}}^o)^T\boldsymbol{w}_k^+ + b_k^+ + 1 - \hat{\eta}_{\hat{j}}\right),
\end{aligned} \tag{4}$$

where $\bar{\alpha}_{\bar{j}}^o \ge 0$, $\hat{\alpha}_{\hat{j}}^o \ge 0$, $\bar{\beta}_{\bar{j}}^o \ge 0$ and $\hat{\beta}_{\hat{j}}^o \ge 0$ are the Lagrange multiplier vectors. Then the Karush-Kuhn-Tucker (KKT) conditions are listed as follows.

$$\frac{\partial L}{\partial \boldsymbol{w}_k^+}=\frac{\bar{S}^k}{\bar{m}^k}\sum_{\bar{i}=1}^{\bar{m}^k}K\varphi(\bar{\boldsymbol{x}}_{\bar{i}}^k)\left(\varphi(\bar{\boldsymbol{x}}_{\bar{i}}^k)^T\boldsymbol{w}_k^+ +b_k^+\right)^2 +\frac{\hat{S}^k}{\hat{m}^k}\sum_{\hat{i}=1}^{\hat{m}^k}\hat{\rho}_{\hat{i}}^k\varphi(\hat{\boldsymbol{x}}_{\hat{i}}^k)\left(\varphi(\hat{\boldsymbol{x}}_{\hat{i}}^k)^T\boldsymbol{w}_k^+ +b_k^+\right) +\sum_{o=k+1,\cdots,N}\sum_{\bar{j}=1}^{\bar{m}^o}\bar{\alpha}_{\bar{j}}^o\varphi(\bar{\boldsymbol{x}}_{\bar{j}}^o)+\sum_{o=k+1,\cdots,N}\sum_{\hat{j}=1}^{\hat{m}^o}\hat{\alpha}_{\hat{j}}^o\varphi(\hat{\boldsymbol{x}}_{\hat{j}}^o)=0, \tag{5}$$

$$\frac{\partial L}{\partial b_k^+}=\frac{\bar{S}^k}{\bar{m}^k}\sum_{\bar{i}=1}^{\bar{m}^k}K\left(\varphi(\bar{\boldsymbol{x}}_{\bar{i}}^k)^T\boldsymbol{w}_k^+ +b_k^+\right) +\frac{\hat{S}^k}{\hat{m}^k}\sum_{\hat{i}=1}^{\hat{m}^k}\hat{\rho}_{\hat{i}}^k\left(\varphi(\hat{\boldsymbol{x}}_{\hat{i}}^k)^T\boldsymbol{w}_k^+ +b_k^+\right) +\sum_{o=k+1,\cdots,N}\sum_{\bar{j}=1}^{\bar{m}^o}\bar{\alpha}_{\bar{j}}^o+\sum_{o=k+1,\cdots,N}\sum_{\hat{j}=1}^{\hat{m}^o}\hat{\alpha}_{\hat{j}}^o=0, \tag{6}$$

$$\frac{\partial L}{\partial \bar{\eta}_{\bar{j}}}=\frac{\bar{C}_1^k K\bar{\Upsilon}^k}{\bar{m}^r}\bar{T}^{k,o}-\bar{\alpha}_{\bar{j}}^o-\bar{\beta}_{\bar{j}}^o=0, \tag{7}$$

$$\frac{\partial L}{\partial \hat{\eta}_j}=\frac{\hat{C}_1^k\hat{\Upsilon}^k}{\hat{m}^r}\hat{T}^{k,o}\hat{\rho}_{\hat{j}}^o-\hat{\alpha}_{\hat{j}}^o-\hat{\beta}_{\hat{j}}^o=0. \tag{8}$$

The combined matrix form of (5) and (6) is

$$\frac{\bar{S}^k K}{\bar{m}^k}\begin{bmatrix}\boldsymbol{A}^T\boldsymbol{A} & \boldsymbol{A}^T\bar{\boldsymbol{e}}_1^k\\ (\bar{\boldsymbol{e}}_1^k)^T\boldsymbol{A} & (\bar{\boldsymbol{e}}_1^k)^T\bar{\boldsymbol{e}}_1^k\end{bmatrix}\begin{bmatrix}\boldsymbol{w}_k^+\\ b_k^+\end{bmatrix} +\frac{\hat{S}^k}{\hat{m}^k}\begin{bmatrix}\boldsymbol{B}^T\hat{\boldsymbol{\rho}}^k\boldsymbol{B} & \boldsymbol{B}^T\hat{\boldsymbol{\rho}}^k\hat{\boldsymbol{e}}_1^k\\ (\hat{\boldsymbol{e}}_1^k)^T\hat{\boldsymbol{\rho}}^k\boldsymbol{B} & (\hat{\boldsymbol{e}}_1^k)^T\hat{\boldsymbol{\rho}}^k\hat{\boldsymbol{e}}_1^k\end{bmatrix}\begin{bmatrix}\boldsymbol{w}_k^+\\ b_k^+\end{bmatrix} =-\sum_{o=k+1,\cdots,N}\begin{bmatrix}\boldsymbol{D}^T\\ (\hat{\boldsymbol{e}}_2^{k,o})^T\end{bmatrix}\hat{\boldsymbol{\alpha}}^o-\sum_{o=k+1,\cdots,N}\begin{bmatrix}\boldsymbol{C}^T\\ (\bar{\boldsymbol{e}}_2^{k,o})^T\end{bmatrix}\bar{\boldsymbol{\alpha}}^o. \tag{9}$$

where

$\boldsymbol{A}=\psi\left(\bar{\boldsymbol{X}}^k,(\tilde{\boldsymbol{X}}^k)^T\right)$, $\boldsymbol{B}=\psi\left(\hat{\boldsymbol{X}}^k,(\tilde{\boldsymbol{X}}^k)^T\right)$, $\boldsymbol{C}=\psi\left(\bar{\boldsymbol{X}}^{k,o},(\tilde{\boldsymbol{X}}^k)^T\right)$

$\boldsymbol{D}=\psi\left(\hat{\boldsymbol{X}}^{k,o},(\tilde{\boldsymbol{X}}^k)^T\right)$, $\bar{\boldsymbol{\alpha}}^o=(\bar{\alpha}_1^o,...,\alpha_{\bar{m}^o}^o)^T$, $\hat{\boldsymbol{\alpha}}^o=(\hat{\alpha}_1^o,...,\bar{\alpha}_{\hat{m}^o}^o)^T$,

$\bar{\boldsymbol{e}}_1^k=[1,...,1]^T$, $\hat{\boldsymbol{e}}_1^k=[1,...,1]^T$, $\bar{\boldsymbol{e}}_2^{k,o}=[1,...,1]^T$, $\hat{\boldsymbol{e}}_2^{k,o}=[1,...,1]^T$,

and $\hat{\boldsymbol{\rho}}^k=\begin{bmatrix}\ddots & 0 & 0\\ 0 & \hat{\rho}_{\hat{i}}^k & 0\\ 0 & 0 & \ddots\end{bmatrix}$,

Then, the solution $\begin{bmatrix}\boldsymbol{w}_k^+\\ b_k^+\end{bmatrix}$ of the hyper-plane h_k^+ in (1) can be obtained from (10):

$$\begin{bmatrix}\boldsymbol{w}_k^+\\ b_k^+\end{bmatrix}=-\left(\begin{bmatrix}\bar{\boldsymbol{H}}^k\\ \hat{\boldsymbol{H}}^k\end{bmatrix}^T\begin{bmatrix}\frac{\bar{S}^k K}{\bar{m}^k}\bar{\boldsymbol{E}}^k & 0\\ 0 & \frac{\hat{S}^k}{\hat{m}^k}\hat{\boldsymbol{\rho}}^k\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{H}}^k\\ \hat{\boldsymbol{H}}^k\end{bmatrix}\right)^{-1}\begin{bmatrix}\vdots\\ \bar{\boldsymbol{G}}^{k,o}\\ \hat{\boldsymbol{G}}^{k,o}\\ \vdots\end{bmatrix}^T\begin{bmatrix}\vdots\\ \bar{\boldsymbol{\alpha}}^o\\ \hat{\boldsymbol{\alpha}}^o\\ \vdots\end{bmatrix}, \tag{10}$$

where $\bar{\boldsymbol{H}}^k=\begin{bmatrix}\boldsymbol{A} & \bar{\boldsymbol{e}}_1^k\end{bmatrix}$, $\hat{\boldsymbol{H}}^k=\begin{bmatrix}\boldsymbol{B} & \hat{\boldsymbol{e}}_1^k\end{bmatrix}$, $\bar{\boldsymbol{G}}^{k,o}=\begin{bmatrix}\boldsymbol{C} & \bar{\boldsymbol{e}}_2^{k,o}\end{bmatrix}$,

$\hat{\boldsymbol{G}}^{k,o}=\begin{bmatrix}\boldsymbol{D} & \hat{\boldsymbol{e}}_2^{k,o}\end{bmatrix}$ and $\bar{\boldsymbol{E}}^k=\begin{bmatrix}\ddots & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & \ddots\end{bmatrix}$.

Based on the KKT conditions, simplify the problem (4), then the dual problem of (2) can be expressed as

$$\min\ \frac{1}{2}\begin{bmatrix}\vdots\\ \bar{\boldsymbol{\alpha}}^o\\ \hat{\boldsymbol{\alpha}}^o\\ \vdots\end{bmatrix}^T\begin{bmatrix}\vdots\\ \bar{\boldsymbol{G}}^{k,o}\\ \hat{\boldsymbol{G}}^{k,o}\\ \vdots\end{bmatrix}\left(\begin{bmatrix}\bar{\boldsymbol{H}}^k\\ \hat{\boldsymbol{H}}^k\end{bmatrix}^T\begin{bmatrix}\frac{\bar{S}^k K}{\bar{m}^k}\bar{\boldsymbol{E}}^k & 0\\ 0 & \frac{\hat{S}^k}{\hat{m}^k}\hat{\boldsymbol{\rho}}^k\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{H}}^k\\ \hat{\boldsymbol{H}}^k\end{bmatrix}\right)^{-1T}\begin{bmatrix}\vdots\\ \bar{\boldsymbol{G}}^{k,o}\\ \hat{\boldsymbol{G}}^{k,o}\\ \vdots\end{bmatrix}^T\begin{bmatrix}\vdots\\ \bar{\boldsymbol{\alpha}}^o\\ \hat{\boldsymbol{\alpha}}^o\\ \vdots\end{bmatrix}-\begin{bmatrix}\vdots\\ \bar{\boldsymbol{e}}_2^{k,o}\\ \hat{\boldsymbol{e}}_2^{k,o}\\ \vdots\end{bmatrix}^T\begin{bmatrix}\vdots\\ \bar{\boldsymbol{\alpha}}^o\\ \hat{\boldsymbol{\alpha}}^o\\ \vdots\end{bmatrix}$$

$$s.t.\quad 0\le\begin{bmatrix}\vdots\\ \bar{\boldsymbol{\alpha}}^o\\ \hat{\boldsymbol{\alpha}}^o\\ \vdots\end{bmatrix}\le\begin{bmatrix}\vdots\\ \frac{\bar{C}_1^k\bar{\Upsilon}^k}{\bar{m}^r}(\bar{\boldsymbol{T}}^o)^T\bar{\boldsymbol{e}}_2^{k,o}\\ \frac{\hat{C}_1^k\hat{\Upsilon}^k}{\hat{m}^r}(\hat{\boldsymbol{T}}^o\hat{\boldsymbol{\rho}}^o)^T\hat{\boldsymbol{e}}_2^{k,o}\\ \vdots\end{bmatrix},\ o=k+1,\ ...,\ N. \tag{11}$$

where $\hat{\boldsymbol{\rho}}^o=\begin{bmatrix}\ddots & 0 & 0\\ 0 & \hat{\rho}_{\hat{j}}^o & 0\\ 0 & 0 & \ddots\end{bmatrix}$, $\bar{\boldsymbol{T}}^o=\begin{bmatrix}\ddots & 0 & 0\\ 0 & \bar{T}^{k,o} & 0\\ 0 & 0 & \ddots\end{bmatrix}$ and

$\hat{\boldsymbol{T}}^o=\begin{bmatrix}\ddots & 0 & 0\\ 0 & \hat{T}^{k,o} & 0\\ 0 & 0 & \ddots\end{bmatrix}$.

Similarly, the dual form of QPP (3) for MHTSVM-*k* sub-model is given as

$$\min\frac{1}{2}\begin{bmatrix}\bar{\boldsymbol{\alpha}}^k\\ \hat{\boldsymbol{\alpha}}^k\end{bmatrix}^T\begin{bmatrix}\bar{\boldsymbol{H}}^k\\ \hat{\boldsymbol{H}}^k\end{bmatrix}\left(\begin{bmatrix}\vdots\\ \bar{\boldsymbol{G}}^{k,o}\\ \hat{\boldsymbol{G}}^{k,o}\\ \vdots\end{bmatrix}^T\begin{bmatrix}\frac{\bar{\Upsilon}^k K}{\bar{m}^r}\bar{\boldsymbol{E}}^{k,o} & 0\\ 0 & \frac{\hat{\Upsilon}^k}{\hat{m}^r}\hat{\boldsymbol{T}}^o\hat{\boldsymbol{\rho}}^o\end{bmatrix}\begin{bmatrix}\vdots\\ \bar{\boldsymbol{G}}^{k,o}\\ \hat{\boldsymbol{G}}^{k,o}\\ \vdots\end{bmatrix}\right)^{-1T}\begin{bmatrix}\bar{\boldsymbol{H}}^k\\ \hat{\boldsymbol{H}}^k\end{bmatrix}^T\begin{bmatrix}\bar{\boldsymbol{\alpha}}^k\\ \hat{\boldsymbol{\alpha}}^k\end{bmatrix}-\begin{bmatrix}\bar{\boldsymbol{e}}_1^k\\ \hat{\boldsymbol{e}}_1^k\end{bmatrix}^T\begin{bmatrix}\bar{\boldsymbol{\alpha}}^k\\ \hat{\boldsymbol{\alpha}}^k\end{bmatrix}$$

$$s.t.\ 0\le\begin{bmatrix}\bar{\boldsymbol{\alpha}}^k\\ \hat{\boldsymbol{\alpha}}^k\end{bmatrix}\le\begin{bmatrix}\frac{\bar{C}_2^k\bar{S}^k}{\bar{m}^k}K\bar{\boldsymbol{e}}_1^k\\ \frac{\hat{C}_2^k\hat{S}^k}{\hat{m}^k}\hat{\boldsymbol{\rho}}^k\hat{\boldsymbol{e}}_1^k\end{bmatrix}, \tag{12}$$

where $\bar{\boldsymbol{\alpha}}^k=(\bar{\alpha}_1^k,...,\alpha_{\bar{m}^k}^k)^T$, $\hat{\boldsymbol{\alpha}}^k=(\hat{\alpha}_1^k,...,\bar{\alpha}_{\hat{m}^k}^k)^T$ and

$\bar{\boldsymbol{E}}^{k,o}=\begin{bmatrix}\ddots & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & \ddots\end{bmatrix}$.

And the solution $\begin{bmatrix}\boldsymbol{w}_k^-\\ b_k^-\end{bmatrix}$ of the hyper-plane h_k^- in (1) is defined as

$$\begin{bmatrix}\boldsymbol{w}_k^-\\ b_k^-\end{bmatrix}=\left(\begin{bmatrix}\vdots\\ \bar{\boldsymbol{G}}^{k,o}\\ \hat{\boldsymbol{G}}^{k,o}\\ \vdots\end{bmatrix}^T\begin{bmatrix}\frac{\bar{\Upsilon}^k K}{\bar{m}^r}\bar{\boldsymbol{E}}^{k,o} & 0\\ 0 & \frac{\hat{\Upsilon}^k}{\hat{m}^r}\hat{\boldsymbol{T}}^o\hat{\boldsymbol{\rho}}^o\end{bmatrix}\begin{bmatrix}\vdots\\ \bar{\boldsymbol{G}}^{k,o}\\ \hat{\boldsymbol{G}}^{k,o}\\ \vdots\end{bmatrix}\right)^{-1}\begin{bmatrix}\bar{\boldsymbol{H}}^k\\ \hat{\boldsymbol{H}}^k\end{bmatrix}^T\begin{bmatrix}\bar{\boldsymbol{\alpha}}^k\\ \hat{\boldsymbol{\alpha}}^k\end{bmatrix}. \tag{13}$$

where

$\bar{\boldsymbol{A}}=\psi\left(\bar{\boldsymbol{X}}^{k,o},(\tilde{\boldsymbol{X}}^k)^T\right)$, $\bar{\boldsymbol{B}}=\psi\left(\hat{\boldsymbol{X}}^{k,o},(\tilde{\boldsymbol{X}}^k)^T\right)$, $\bar{\boldsymbol{C}}=\psi\left(\bar{\boldsymbol{X}}^k,(\tilde{\boldsymbol{X}}^k)^T\right)$

and $\bar{\boldsymbol{D}} = \psi\left(\hat{\boldsymbol{X}}^k, (\tilde{\boldsymbol{X}}^k)^T\right)$.

IV. Experiments

In the section, the proposed MHTSVM model is verified on six types of defects datasets collected from different steel production line with the vision-based acquisition system. More details of these defect samples have been described in section II.A. In order to verify the effectiveness and adaption of MHTSVM model, a series of experiments are carried out. In order to obtain better results, the grid search method is used to find optimal parameters. For MHTSVM model, the following sets of parameters are explored: the penalty parameters $\bar{C}_1^k$, $\hat{C}_1^k$, $\bar{C}_2^k$ and $\hat{C}_2^k$ are chosen from $[2^{-7}, 2^7]$, the kernel radius δ is searched over the set $[2^{-7}, 2^7]$, the parameter K is set as 8. All algorithms are implemented in MATLAB 2016b [21]. All algorithms are operated on Windows 10 running on a PC with an Intel I5 processor and 8 GB RAM.

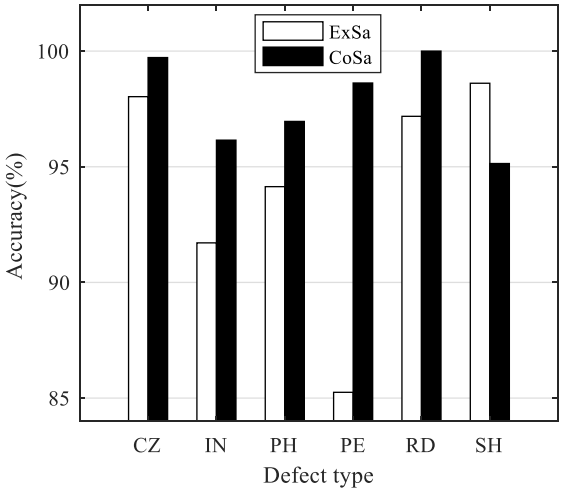


Fig. 2. The comparison of classification accuracy for expert samples and comprehensive samples.

TABLE II
The classification accuracy of expert samples and comprehensive samples

Defect type	ExSa (A0) (%)	CoSa (A1) (%)	A1-A0 (%)
CZ	98.03	**99.72**	1.69
IN	91.71	**96.15**	4.44
PH	94.14	**96.96**	2.82
PE	85.25	**98.62**	13.37
RD	97.18	**100**	2.82
SH	**98.61**	95.14	-3.47
Total acc.	94.05	**97.77**	3.72

Firstly, in order to verify the rationality of the existence of rough samples, a group of comparative experiments are carried out for expert samples and comprehensive samples (CoSa). The comprehensive samples are composed of expert samples and rough samples. In the experiment, the MHTSVM^0 model is trained. In MHTSVM^0, both the gaussian weight information and within-class scatter information are canceled. In fact, MHTSVM^0 is equal to TSVM model. The testing results are shown in Table II and compared in Fig. 2. From the results, it can be seen that the total accuracy (acc.) of CoSa is higher than that of ExSa for 3.72%. The reason is that the size of expert samples is insufficient. If the sample size is insufficient, the results we learn from the model may be wrong, because there will be many rules in the case of insufficient samples. On the contrary, there is a large amount of rough samples, which can reflect specific rule, and the result is naturally improved. On the other hand, Table II further show the classification accuracies for each type of defects. The best accuracy is bolded. The difference in accuracy between CoSa and ExSa is calculated. It can be seen that the classification accuracies obtained by comprehensive samples reach to the best for most types of defects. In short, this indicates that the introduction of rough samples is reasonable. Therefore, all the next experiments are performed on the comprehensive samples.

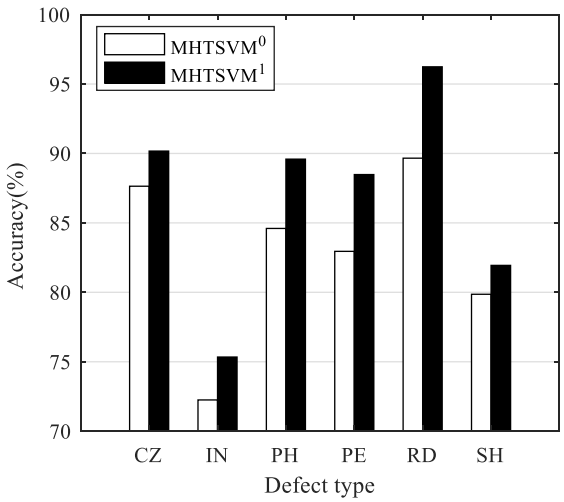


Fig. 3. The comparison of classification accuracy for two classifiers in corrupted defects dataset.

TABLE III
The classification accuracy of two classifiers in corrupted defects dataset

Defect type	MHTSVM^0 (A0) (%)	MHTSVM^1 (A1) (%)	A1-A0 (%)
CZ	87.64	**90.17**	2.53
IN	72.25	**75.34**	3.09
PH	84.60	**89.59**	4.99
PE	82.95	**88.48**	5.53
RD	89.66	**96.24**	6.58
SH	79.86	**81.94**	2.08
Total acc.	82.24	**86.41**	4.17

Secondly, the experiments are done to testify the performance of gaussian weight information in MHTSVM model. In reality, there are inevitably noise in rough samples. These noise will affect classification results. The guassian weight was used to restrain the adverse effect of noise in MHTSVM model. For better illustration, a corrupted defects dataset is built by introducing label noise into the original defects dataset, where the ratio between the number of noise samples and that of training samples is 5%. Firstly, MHTSVM^0 and MHTSVM^1 models are trained with the corrupted dataset. MHTSVM^1 is a simplified version of MHTSVM. In MHTSVM^1, only gaussian weight information is considered. The within-class scatter information is ignored in MHTSVM^1. Then, the original testing dataset is adopted to test the precision for different types of defects. The testing

results of two models are recorded in Table III and compared in Fig. 3. It can be seen that the total accuracy of MHTSVM[1] is higher than that of MHTSVM[0] for 4.17%. Obviously, the MHTSVM[1] model using $\rho(\hat{\boldsymbol{x}}_i^k)$ yields the best classification accuracy. On one hand, the gaussian weight information can strengthen the effect of important samples. On the other hand, the weight information can suppress the adverse effect of noise samples on the classification model. Hence, for the corrupted dataset, the gaussian weight information enhances the classification accuracy of MHTSVM[1]. The result fully proves the superiority of the gaussian weight in MHTSVM model.

Thirdly, the feasibility of within-class scatter information in MHTSVM model is verified. In this experiment, two models are tested in original defects dataset. One is simplified MHTSVM[2] which only uses the within-class scatter information. And the other is MHTSVM[0] model. To evaluate the performance of the MHTSVM[2] model, the evaluation criteria should be specified. Assume that the specificity is TPR, and the sensitivity is TNR, which can be calculated as follows:

$$TPR = TP/(TP+FN), \tag{14}$$

$$TNR = TN/(TN+FP), \tag{15}$$

where TP is the number of true positive samples, FP is the number of false positive samples, TN is the number of true negative samples, and FN is the number of false negative samples. Then, the following criteria can be defined.

$$G = \sqrt{TPR \times TNR}, \tag{16}$$

$$Perr = 1 - TPR, \tag{17}$$

$$Nerr = 1 - TNR, \tag{18}$$

$$Merr = 1 - [(1-Perr)(1-Nerr)]^{1/2}. \tag{19}$$

TABLE IV
THE RESULTS OF MHTSVM[0] AND MHTSVM[2] FOR UNBALANCED CLASSIFICATION

k	Error-rate	MHTSVM[0]-k	MHTSVM[2]-k
1	*Perr*	0.0056	0.0028
	Nerr	0.0006	0.0012
	Merr	0.0031	0.0020
	G	0.9969	**0.9980**
2	*Perr*	0.0347	0.0347
	Nerr	0.0175	0.0175
	Merr	0.0261	0.0262
	G	**0.9739**	0.9738
3	*Perr*	0.0022	0.0022
	Nerr	0.0231	0.0231
	Merr	0.0127	0.0127
	G	**0.9873**	**0.9873**
4	*Perr*	0.0140	0.0093
	Nerr	0.0043	0.0043
	Merr	0.0091	0.0068
	G	0.9909	**0.9932**
5	*Perr*	0	0
	Nerr	0.0137	0.0069
	Merr	0.0069	0.0035
	G	0.9931	**0.9965**

In the above equations, G is the G-mean, $Perr$ the misclassification rate of target samples, $Nerr$ is the misclassification rate of background samples, and $Merr$ the average classification rate. G-mean is the geometric evaluation of sensitivity and specificity. When the values of sensitivity and specificity are high, the value of G-mean will be large. Compared with specificity and sensitivity, G-mean can more accurately describe the performance of classifier. Generally, a greater G value indicates that MHTSVM[2] model has a better performance. The final testing results are listed in Table IV. It can be observed from Table IV that the G values of MHTSVM[2] on four sub-models are optimal, except for the second sub-model. In addition, it can be seen that the values of $Perr$ and $Merr$ can be reduced under the premise of keeping the $Nerr$ unchanged for MHTSVM[2]. This result indicates that the MHTSVM[2] model using within-class scatter information can better deal with unbalanced dataset.

Fouthly, in order to highlight the efficacy of the proposed multi-class classification method, MHTSVM model is compared with SVM, TSVM and KNN on the corrupted defects dataset mentioned in the second experiment. In order to intuitively show the effectiveness of the classification results, the confusion matrix is used to evaluate the test results in the experiment, as shown in Fig. 4. The correct classification results of each type of defect are recorded in Table V. It can be seen that our MHTSVM obtains the best classification performance on all four defects except for the "PH" and "RD" defect datasets. In addition, although the accuracy of MHTSVM on the other two datasets is not optimal, the MHTSVM is better than TSVM. This further verifies the feasibility and applicability of MHTSVM algorithm.

In addition, Fig. 5 records the total accuracies of four algorithms. As can be observed from Fig. 5, MHTSVM is at least 4% higher than other three classifiers in the total accuracy. The reason is that MHTSVM includes the gaussian weight information. A small weight shows that the sample has little influence on training results. It can reduce the adverse impact of noise samples. This suggests that the proposed MHTSVM plays an anti-noise role, while other three classifiers are affected by noise samples. The other reason is that MHTSVM adopts within-class scatter information which is helpful to improve the classification performance of unbalanced data and enhance the classification accuracy.

Lastly, the statistical analysis methods are used to further illustrate the superiority of our MHTSVM model. Firstly, based on the Table V, the classification accuracies of four classifiers are ranked, and the results are recorded in Table VI. The lower the rank is, the better the performance of the classifier is. Obviously, the average rank of MHTSVM is the smallest among the four classifiers. This shows that MHTSVM is more effective than other three classifiers.

Assuming that all algorithms are equivalent, then the Friedman statistic parameter [22] can be calculated using the average rank, which is defined as

$$\chi_F^2 = \frac{12l}{m(m+1)}\left[\sum_j Rank_j^2 - \frac{m(m+1)^2}{4}\right], \tag{20}$$

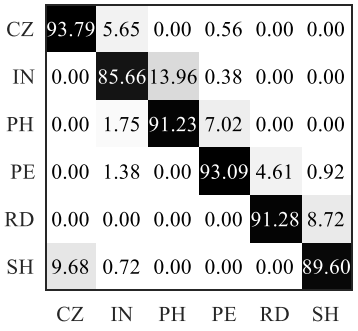
where $Rank_j = \frac{1}{l}\sum_{i=1}^{l} rank_i^j$ and $rank_i^j$ refers to the jth of m algorithms on the ith of l datasets. Friedman's χ_F^2 is so conservative that a better statistic is produced.

$$F_F = \frac{(l-1)\chi_F^2}{l(m-1)-\chi_F^2}, \tag{21}$$

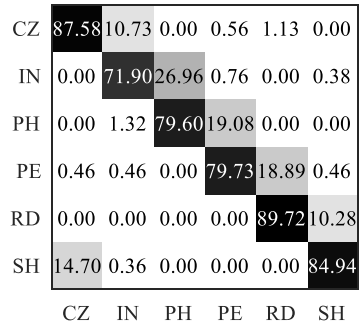
which is distributed according to F-distribution with (m-1) and (m-1) (l-1) degrees of freedom. Table VII lists the result of the Friedman test. When the level of significance $\alpha = 0.05$, according to (20) and (21), the critical value of $F(3,15)$ is 3.287. When $\alpha = 0.1$, the critical value of $F(3,15)$ is 2.490. Because $F_F = 5.485$ which is larger than 3.287 and 2.490, α is less than 0.05. As can be seen from Table VII, the value F_F is far greater than the critical value F_α. Therefore, we reject the null-hypothesis, which means that there are significant differences between the four classifiers.

TABLE V
THE CLASSIFICATION ACCURACY OF FOUR CLASSIFIERS IN CORRUPTED DEFECTS DATASET

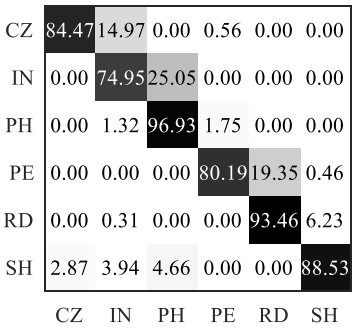
Defect type	KNN (%)	SVM (%)	TSVM (%)	MHTSVM (%)
CZ	88.14	84.47	87.58	**93.79**
IN	75.53	74.95	71.90	**85.66**
PH	79.38	**96.93**	79.60	91.23
PE	77.88	80.19	79.73	**93.09**
RD	88.17	**93.46**	89.72	91.28
SH	68.09	88.53	84.94	**89.60**



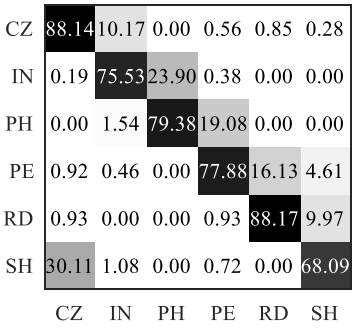
(a) MHTSVM



(b) TSVM



(c) SVM



(d) KNN

Fig. 4. The confusion matrix.

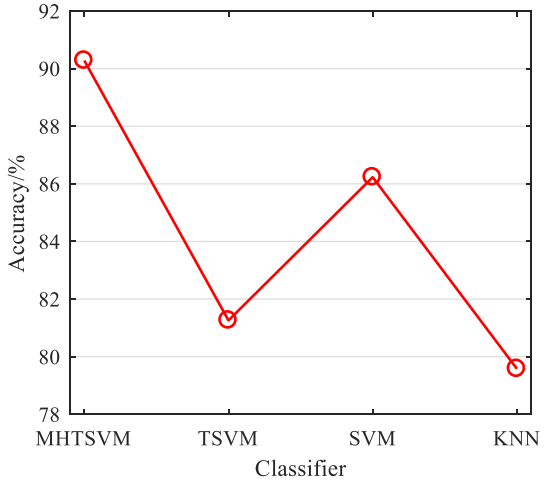


Fig. 5. The total accuracies of four classifiers in corrupted defects dataset.

TABLE VI
AVERAGE RANK OF FOUR CLASSIFIERS IN TERMS OF CLASSIFICATION ACCURACY IN TABLE V

Defect type	KNN	SVM	TSVM	MHTSVM
CZ	2	4	3	1
IN	2	3	4	1
PH	4	1	3	2
PE	4	2	3	1
RD	4	1	3	2
SH	4	2	3	1
Average Rank	3.33	2.17	3.17	**1.33**

Next, to judge the significant difference between two algorithms, the Bonferroni-Dunn test [23] is adopted. The critical difference (CD) of the Friedman test with Bonferroni-Dunn test is calculated by (22). The results are listed in Table VII.

$$\mathrm{CD}_{(\alpha=0.1)} = q_\alpha \sqrt{\frac{m(m+1)}{6\times l}}. \tag{22}$$

If the difference between the ranks of two algorithms is larger than CD, their performance is considered to be significantly different. From Table VII, we can have that when $\alpha = 0.05$, from (22), we have the critical difference CD is 1.91. When $\alpha = 0.1$, we have CD=1.59. Obviously, when $\alpha = 0.05$, only the differences between the ranks of our MHTSVM and KNN is large than 1.91. But, when $\alpha = 0.1$, we can conclude that except for SVMs model, the differences between the ranks of our MHTSVM and other two algorithms are large than 1.59. This further indicates that the proposed MHTSVM model shows great advantage in the steel surface defects classification.

TABLE VII
THE RESULTS IN THE FRIEDMAN TEST AND THE BONFERRONI-DUNN TEST

Significance	χ_F^2	F_F	F_α	q_α	CD
$\alpha = 0.05$	9.416	5.485	3.287	2.569	1.91
$\alpha = 0.1$	9.416	5.485	2.490	2.291	1.59

V. CONCLUSIONS

In this paper, we propose a new multi-class classification model for steel surface defects, which is termed as MHTSVM. MHTSVM can be split into three sub-models: MHTSVM^0, MHTSVM^1 and MHTSVM^2. MHTSVM^0 formulates TSVM-type classification model, which makes MHTSVM^0 have good generalization performance. Differently, MHTSVM^0 is trained by learning expert samples and rough samples respectively, which makes MHTSVM^0 more adaptable. The superiority of MHTSVM^0 is proved by comparative experiments with expert samples and rough samples. MHTSVM^1 which is based on MHTSVM^0 adopts KNN method to describe the gaussian weight information in the training samples. It not only suppresses noise samples, but also strengthens important samples. Both robustness and reliability of MHTSVM^1 are proved by comparing MHTSVM^0 and MHTSVM^1 with experiments. MHTSVM^2 introduces the within-class scatter information, which effectively improves the classification performance on unbalanced datasets. The performance of MHTSVM^2 has been verified by experiments. Subsequently, combining the above three sub-models, the novel MHTSVM is established. MHTSVM is not only insensitive to noise, but also can effectively solve the classification problem on unbalanced datasets. Finally, the novel MHTSVM is compared with SVM, TSVM and KNN by a series of experiments. The results verify that MHTSVM obtains a very satisfactory success rate, especially for corrupted defect samples.

REFERENCES

[1] F. Pernkopf, "Detection of surface defects on raw steel blocks using Bayesian network classifiers," Pattern Analysis and Applications, vol. 7, no. 3, pp333-342, 2004.

[2] Q. Luo and Y. He, "A cost-effective and automatic surface defect inspection system for hot-rolled flat steel," Robotics and Computer-Integrated Manufacturing, vol. 38, pp16-30, 2016.

[3] M. Jiang, G. Li, L. Xie, M. Xiao and L. Yi, "Adaptive classifier for steel strip surface defects," Journal of Physics: Conference Series. IOP Publishing, vol. 787, no. 1, pp012019, 2017.

[4] F. Dupont, C. Odet and Cartont M, "Optimization of the recognition of defects in flat steel products with the cost matrices theory," NDT & E International, vol. 30, no.1, pp3-10, 1997.

[5] E. Amid, S. R. Aghdam and H. Amindavar, "Enhanced performance for support vector machines as multiclass classifiers in steel surface defect detection," International Journal of Electrical and Computer Engineering, vol. 6, no. 7, pp693-697, 2012.

[6] C. S. Lee, C. H. Choi, J. Y. Choi, Y. K. Kim and S. H. Choi, "Feature extraction algorithm based on adaptive wavelet packet for surface defect classification," Proceedings of 3rd IEEE International Conference on Image Processing. IEEE, vol. 2, pp673-676, 1996.

[7] L. Yi, G. Li and M. Jiang, "An end-to-end steel strip surface defects recognition system based on convolutional neural networks," steel research international, vol. 88, no. 2, pp1600068, 2017.

[8] D. Cui and K. Xia, "Strip surface defects recognition based on PSO-RS&SOCP-SVM algorithm," Mathematical Problems in Engineering, vol. 2017, pp1-9, 2017.

[9] L. Liu, M. Chu, R. Gong and X. Qi, "Unbalanced classification method using least squares support vector machine with sparse strategy for steel surface defects with label noise," Journal of Iron and Steel Research International, vol. 27, no. 12, pp1407-1419, 2020.

[10] Vapnik V, *The nature of statistical learning theory* (Book style). Springer Science & Business Media, 1999.

[11] N. Neogi, D. K. Mohanta and Dutta P K, "Review of vision-based steel surface inspection systems," EURASIP Journal on Image and Video Processing, vol. 2014, no. 1, pp1-19, 2014.

[12] N. Ketkar, *Convolutional Neural Networks* (Book style). Springer International Publishing, 2017.

[13] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp273-297, 1995.

[14] Md. Abdul Alim Sheikh, Tanmoy Maity and Alok. Kole, "Deep Learning Approach using Patch-based Deep Belief Network for Road Extraction from Remote Sensing Imagery," IAENG International Journal of Applied Mathematics, vol. 52, no. 4, pp760-775, 2022.

[15] R. Khemchandani and S. Chandra, "Twin support vector machines for pattern classification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 5, pp905-910, 2007.

[16] X. Peng and X. Dong, "Bi-density twin support vector machines for pattern recognition," Neurocomputing, vol. 99, no. JAN.1, pp134–143, 2013.

[17] Z. Xue, S. Liu and W. Liu, "Unbalanced least squares support vector machines," Journal of System Simulation, vol. 21, no. 14, pp4324-4327, 2009.

[18] A. K. Nain, S, Gupta and B. Bhushan, "An extension to switching bilateral filter for mixed noise removal from color image," International Journal of Signal and Imaging Systems Engineering, vol. 9, no. 1, pp1-19, 2016.

[19] L. Y. Zhang and B. A.Rosdi, "Region of interest extraction of finger-vein image using watershed segmentation with distance transform,"//InECCE2019. Springer, Singapore, pp333-345, 2020. DOI: 10.1007/978-981-15-2317-5_28

[20] B Wang, B Lv and Y Song, "A Hybrid Genetic Algorithm with Integer Coding for Task Offloading in Edge-Cloud Cooperative Computing," IAENG International Journal of Computer Science, vol. 49, no. 2, pp503-510, 2022.

[21] The Math Works (MATLAB 2016b), Inc. [Online]. Available: http://www.mathworks.com.

[22] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," The Journal of Machine Learning Research, vol. 7, pp1-30, 2006.

[23] D. G. Pereira, A. Afonso and F. M. Medeiros. Overview of Friedman's test and post-hoc analysis[J]. Communications in Statistics-Simulation and Computation, vol. 44, no. 10, pp2636-2653, 2015.

MAOXIANG CHU was born in September 7, 1978. He is an associate professor and doctoral supervisor at School of Electronic and Information Engineering in University of Science and Technology Liaoning. His research interests include pattern recognition, machine learning, image processing and intelligent control.

ZIXUAN ZHAI was born in May 29, 1997. She is a graduate student at School of Electronic and Information Engineering in University of Science and Technology Liaoning. Her main research direction is pattern recognition and machine learning.

LIMING LIU was born in June 10, 1994. She is a Ph.D. student at School of Electronic and Information Engineering in University of Science and Technology Liaoning. Her main research direction is pattern recognition and machine learning.

GUANGHU LIU was born in October 4, 1998. He is a Ph.D. student at School of Electronic and Information Engineering in University of Science and Technology Liaoning. His main research direction is pattern recognition and machine learning.